SYSTEMS BIOLOGY OF THERMOTOGA NEAPOLITANA FOR BIOLOGICAL HYDROGEN PRODUCTION

by Sarah Ann Munro

SYSTEMS BIOLOGY OF *THERMOTOGA NEAPOLITANA* FOR BIOLOGICAL HYDROGEN PRODUCTION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sarah Ann Munro

August 2010

SYSTEMS BIOLOGY OF *THERMOTOGA NEAPOLITANA* FOR BIOLOGICAL HYDROGEN PRODUCTION

Sarah Ann Munro, Ph. D.

Cornell University 2010

The hyperthermophile *Thermotoga neapolitana* was evaluated for hydrogen production in three phases of iterative experimental and computational modeling work. First, the effect of environmental perturbations and fermentation stoichiometry were experimentally determined. Variation of growth temperature within the permissible range did not affect product yields. Maximum H_2 production rates were achieved in both 77 and 85 °C experiments. Low pH inhibited glucose consumption; when pH was raised, glucose consumption was complete. Oxygen exposure tests confirmed that O_2 addition did not increase H_2 production. The fermentation balance at 85 °C was 2.8 mol H_2, 2 mol CO_2, 1.8 mol acetate, and 0.1 mol lactate per mol of glucose consumed.

Constraint-based analysis can increase understanding of metabolic network interactions and can predict effects of metabolic engineering thereby decreasing the need for costly experiments. A comparative reconstruction method was created to convert a constraint-based model for *Thermotoga maritima* into a model for *T. neapolitana* based on synteny between the two annotated genome sequences. Flux balance analysis simulations of *T. neapolitana* batch growth were validated with previously obtained experimental results. The *T. neapolitana* model was examined to identify mechanisms to allow growth with cysteine as a sole sulfur source, because unlike *T. maritima* which can only use elemental sulfur in a defined medium, *T. neapolitana* can also use cysteine. The results were inconclusive; all genes with

known functions related to sulfur metabolism were found in both species. Recent evidence in the literature that both species have a bifurcating hydrogenase that simultaneously utilizes ferredoxin and NADH was also incorporated into the *T. neapolitana* model. The modified model required the inclusion of a membrane bound NADH:Ferredoxin oxidoreductase to maintain the correct NAD^+/NADH ratio to support growth.

Experimental validation of *T. neapolitana* model-derived hypotheses was conducted. Initial carbon substrate utilization predictions indicated that glycerol, L-rhamnose, and cellotetraose could not support growth. Experimental results showed that both glycerol and L-rhamnose did not support growth, however cellotetraose did support growth. Further analysis comparing protein expression data from cellotetraose and glucose grown cells suggests that the model can be updated to include a more complete cellotetraose pathway.

BIOGRAPHICAL SKETCH

Sarah Munro was born July 11, 1980 in Buffalo, NY to Mary Ellen and Kevin Munro. She graduated from East Aurora High School in East Aurora, NY in 1998. Sarah's interest in the engineering application of biology to address and prevent environmental problems led her to enter Cornell University to study Biological and Environmental Engineering. After earning her Bachelor of Science degree in Biological and Environmental Engineering, she chose to continue her education at Cornell in pursuit of a Doctorate of Philosophy Degree under the guidance of Dr. Larry Walker.

Sarah plans to move to the Washington D.C. /Baltimore, MD area where she will pursue research opportunities at the intersection of experimental and computational biology.

For my mother and father

ACKNOWLEDGMENTS

I am grateful to the many people who have supported me professionally and personally during my time at Cornell. I hope that anyone I have forgotten to acknowledge here understands my gratitude nonetheless.

First I would like to thank my major professor Dr. Larry Walker. I have known Larry since 2000, and worked with him as an undergraduate and graduate student. Dr. Walker has been and will continue to be my most important mentor. He has consistently offered me excellent advice and opportunities that I will always be grateful for. I hope to be as great a mentor to others as Larry has been to me.

Sincerest thanks go to my other committee members, Dr. Stephen Zinder, Dr. James Gossett, and Dr. David Wilson. All three have been excellent advisors and I appreciate all the insight and advice they've given to me in our meetings. I've worked most closely on my project with Dr. Zinder, and I am grateful to him for sharing his expertise with me. Dr. Gossett and Dr. Wilson have both offered significant support by giving me access to their laboratories and encouraging me to connect with other researchers (their students and research associates and most importantly Dr. Zinder) so that I could address my research challenges.

Many other researchers have been immensely helpful to me during my project. Dr. Kelvin Lee has been an important collaborator for proteomic analysis. I am grateful for the patient training provided by Diana Irwin. I would also like to thank some fellow graduate students for their help with my project: Rodrigo Labatut for allowing me to use his pressure transducer system, Iman Elgheriany for my initial training on the total organic carbon analyzer, and Ben Heavner for taking the plunge into systems biology with me.

I am grateful for all of Sue Fredenburg's help over the years; she has always provided great support for the Walker lab graduate students. I'm grateful to Dr. Stéphane Corgié for thoughtful research discussions. I also would like to thank Dr. Corey Rutzke, Dr. Jose Moran-Mirabal, and Ed Evans for their extensive efforts to keep our research program running.

During my time at Cornell I've worked and relaxed with some great people including: Aaron Saathoff, Ben Heavner, Caroline Corner, Deborah Sills, Dong Yang, Erin Brown, Hnin Aung, Jeremy Luterbacher, John Connelly, Kate O'Connor, Kristy Graf, Linelle Fontenelle, Marie Donnelly, Melissa Brechner, Mimi Perez-Falcon, Navaneetha Santhanam, Scott Pryor, and Sonya Padron. I've valued all the time I've spent with each of you.

I wouldn't be where I am today without all the love and support I've received from my family: my parents, Mary Ellen and Kevin, grandparents, Buster and Loretta, and younger brother, Brian. My mother in particular has been incredibly important with her boundless moral support. I am thankful to her for all she has given and shown me, and for suggesting Cornell in the first place.

Last, but not least, I want to thank my fiancé, Lucas Koerner. Having him by my side during graduate school has made every day a little bit better. I am grateful to him for being my editor and best friend.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Although presently there is a major R&D emphasis on liquid fuel replacements for gasoline, in the future energy carriers such as hydrogen (H_2) will play a significant role in meeting our energy needs. Hydrogen is generated on an industrial-scale from non-renewable resources; however as market demands increase and fossil fuels are depleted additional methods of H_2 production based upon renewable resources such as biomass will become crucial. One such method, biological hydrogen production, employs microorganisms to produce H_2. One of the most promising species is the hyperthermophilic bacterium, *Thermotoga neapolitana*, which converts sugars to H_2 via fermentation. Growth at high temperatures gives advantages to *T. neapolitana*, the risk of contamination is lower and the possible yields are higher than those that are achievable with mesophilic systems. The use of a pure culture can be advantageous over mixed communities of unknown composition, because the metabolism of the organism can be specifically modified and controlled to improve H_2 yields.

There is a large amount of information regarding *T. neapolitana* carbohydrate metabolism in the literature. Although some progress has been made in understanding metabolism in *T. neapolitana*, the lack of a robust genetic system and, until recently, an annotated genome has limited the depth of understanding in this microorganism. Thus, most of the reported literature on *Thermotoga* species has been on functional genomics research on *Thermotoga maritima*, a closely related species.

Metabolic modeling strategies have only recently been applied to *T. maritima* and the development of a predictive *in silico* model for *T. neapolitana* would be useful not only for improving H_2 production yields in this species, but also for advancing our

understanding of metabolism in *Thermotogales* species. Most mathematical models of H_2 fermentation are limited to unstructured kinetic and statistical models; the advantage of a structured metabolic model is that constraint-based modeling can be employed to predict the response of an organism to environmental and genetic perturbations and generate new hypotheses.

The goal of this research was to use an iterative systems biology approach of constraint-based modeling to study H_2 production in *T. neapolitana*. First, experimental analysis was conducted to determine the *T. neapolitana* fermentation stoichiometry and response to environmental perturbations. In the second phase of this dissertation a constraint-based *in silico* model of *T. neapolitana* central carbon metabolism was built. A new comparative model reconstruction method was created to convert a constraint-based model for *Thermotoga maritima* into a model for *T. neapolitana* based on synteny between the two annotated genome sequences. Similar to comparative genomics, this comparative model development method is an expeditious procedure to generate a model by extending the utility of a related model and bypassing the detailed database searching and integration required for the original model reconstruction process. Experimental data from the first physiological study was used to test the accuracy of the model. In the final phase of this dissertation, additional experiments were conducted to test hypotheses generated from the *T. neapolitana* model. This experimental validation of the model predictions is critical to improve the accuracy of the model for future computational hypothesis generation.

Chapter 2

Literature Review

2.1 Biological Hydrogen (H_2) Production

There is considerable national interest in developing biological H_2 production technologies, because current H_2 production technologies rely on nonrenewable resources, including steam reforming of methane in natural gas and gasification of coal (Turner 2004). The types of biological H_2 production can be divided into three major categories:

1. Photoautotrophic – Bio-photolysis of water
2. Photoheterotrophic – Photo-fermentation of organic matter
3. Heterotrophic – Fermentation (non-photolytic or dark) of organic matter

Photoautotrophic and photoheterotrophic organisms both rely on light energy captured by photosynthesis to drive metabolism and H_2 evolution. However, photoautotrophic organisms use CO_2 as a carbon source, whereas photoheterotrophic organisms must use organic carbon. Several species of algae and cyanobacteria are capable of photoautotrophic H_2 production. The overall reaction for photoautotrophic H_2 production is also referred to as bio-photolysis because light energy harvested from photosynthesis is used to split H_2O to form O_2 and H_2 as in the reaction:

$$2H_2O \xrightarrow{\text{light}} O_2 + 2H_2 \qquad (2.1)$$

More specifically, electrons that are generated from photosynthesis are used by hydrogenase enzymes to reduce protons to molecular H_2 according to the reaction:

$$2H^+ + 2e^- \rightarrow H_2 \qquad (2.2)$$

Chlamydomonas reinhardtii is the most extensively studied H_2 producing algal species (Flynn et al. 2002; Tsygankov et al. 2002), but other species include *Scenedesmus obliquus, Chlorococcum littorale, Platymonas subcordiformis,* and *Chlorella fusca* (Kapdan and Kargi 2006).

Cyanobacteria also use light energy from photosynthesis to drive H_2 production, but employ an indirect method of bio-photolysis. Rather than splitting H_2O directly to form H_2, intermediate compounds are first formed according to the generalized photosynthetic reaction:

$$nH_2O + nCO_2 \xrightarrow{\text{light}} C_n(H_2O)_n + nO_2 \quad (2.3)$$

The intermediate compounds, typically glucose or organic acids, are then endogenously metabolized using light energy to produce H_2 (Akkerman et al. 2002; Levin et al. 2004):

$$C_n(H_2O)_n + nH_2O \xrightarrow{\text{light}} nCO_2 + 2nH_2 \quad (2.4)$$

Cyanobacterial species under investigation include nitrogen-fixing species such as *Anabaena* sp., *Oscillatoria* sp., *Calothrix* sp., and non-nitrogen fixing organisms including *Synechococcus* sp. and *Gloebacter* sp. (Kapdan and Kargi 2006). *Anabaena variabilis* is the most studied species for cyanobacterial H_2 production, because high production rates have been achieved in some strains (Markov et al. 1997; Borodin et al. 2000; Lopes Pinto et al. 2002; Tsygankov et al. 2002; Levin et al. 2004). Cyanobacteria have a variety of enzymes that directly catalyze reactions involving H_2. These include nitrogenases, which produce H_2 as a co-product of nitrogen fixation, uptake hydrogenases, which oxidize H_2 that is synthesized by the nitrogenase activity, and bidirectional hydrogenases, which are capable of oxidizing and producing H_2 (Hansel and Lindblad 1998; Tamagnini et al. 2002; Levin et al. 2004). Limitations of algal and cyanobacterial systems include inhibition of hydrogenase activity by the O_2 produced during photoautotrophic growth (Benemann 1997; Akkerman et al. 2002;

Hallenbeck and Benemann 2002) and low reported photosynthetic conversion efficiencies, less than 1% of solar energy is converted to H_2, and the theoretical maximum is 10% conversion efficiency (Hallenbeck and Benemann 2002).

Photoheterotrophic H_2 production via photo-fermentation of organic matter is performed by purple photosynthetic bacteria species, including *Rhodobacter spheroides*, *Rhodobacter capsulatus*, *Rhodovulum sulfidophilum*W-1S, and *Rhodopseudomonas palustris* (Kapdan and Kargi 2006). The overall reaction for photo-fermentation using acetate as an example electron donor (Akkerman et al. 2002) is:

$$C_2H_4O_2 + 2H_2O \xrightarrow{\text{light}} 2CO_2 + 4H_2 \qquad (2.5)$$

Nitrogenases are the key enzymes for this reaction system. Hydrogenases are present, but possess H_2 uptake activity that catalyze the reverse reaction of oxidizing molecular H_2 to protons; this is seen as a major limitation to the use of these organisms for H_2 production (Koku et al. 2002; Kapdan and Kargi 2006). Another drawback of this system is that nitrogenase activity has a high ATP requirement for H_2 evolution – 4 ATP per H_2 evolved (Hallenbeck and Benemann 2002).

The third category, heterotrophic fermentation, is appealing for several reasons. First, fermentation is the basis of a variety of industrial processes, ranging from the production of foods and beverages to pharmaceuticals, polymers, and fuels. Existing reactor technology is in place and many of the engineering issues at the industrial scale have been addressed, whereas new reactor designs are required to obtain appropriate light-exposure for photosynthetic organisms (Akkerman et al. 2002; Rupprecht et al. 2006). The lack of light requirement in fermentative organisms is also advantageous, because H_2 can be produced continuously without sunlight or other light sources (Das and Veziroglu 2001). Another important benefit of employing fermentation to produce H_2 is that it can be a waste treatment strategy (Kapdan and

Kargi 2006). Using industrial food processing, agricultural wastes, etc. for production of H_2 not only reduces waste streams, but also extracts value from these materials. Fermentative H_2 production systems will also have the advantage of benefiting from the immense amount of R&D currently focused on the development of low-cost sugars from lignocellulosic materials for ethanol production; the same pretreatment and hydrolysis strategies could be employed to provide low-cost sugars for fermentative H_2 production.

2.1.1 Fermentative H_2 Production

2.1.1.1 Key Pathways and Enzymes

Fermentation is defined as non-photolytic catabolism in which organic carbon compounds are used as both electron donors and acceptors, instead of an external electron acceptor (such as oxygen or nitrate). Organisms that use fermentation are often strict anaerobes and cannot tolerate the presence of oxygen (O_2); however, facultative anaerobes perform fermentation under anaerobic conditions and also aerobic respiration in the presence of O_2.

The metabolic pathway of fermentation can be divided into two major stages. The first stage is glycolysis, the conversion of glucose to the intermediate compound pyruvate. For every 6-carbon glucose molecule that enters the glycolysis pathway two 3-carbon pyruvate molecules are produced. In the second stage of fermentation, pyruvate is catabolized to a variety of oxidized and reduced end-products, which may include acetate, butyrate, lactate, propionate, ethanol, CO_2, and H_2, etc. Greater detail on these pathways and related pathways such as extracellular carbohydrate processing,

membrane transport systems, and the pentose-phosphate pathway is provided in section 2.2.

The ratios of the end-products from fermentation vary between species and also change depending on the type of carbohydrate substrate, the growth phase of the organism, and environmental parameters, such as temperature and pH. For example, for H_2 production from glucose the ideal overall reaction would be complete oxidation of glucose to H_2 and CO_2:

$$C_6H_{12}O_6 + 12H_2O \longrightarrow 6HCO_3^- + 6H^+ + 12H_2 \quad (2.6)$$

However, this ideal conversion has not been observed *in vivo* for any organisms (Westermann et al. 2007), because at standard conditions and pH the change in Gibbs free energy, $\Delta G°'$, for complete oxidation of glucose to H_2 is slightly positive, $\Delta G°' = +3.2$ kJ·mol^{-1} (Thauer et al. 1977). This indicates that energy is required for the reaction and therefore it is not useful for catabolism in which the overall goal is to generate energy that can be used for growth and maintenance of the organism. Reactions with a net negative $\Delta G°'$ are more favorable, however some catabolic reactions can proceed with a positive $\Delta G°'$ if light energy is utilized, but this only occurs in photosynthetic organisms.

The theoretical maximum H_2 yield achievable through fermentation by whole organisms is obtained when acetate is the primary volatile fatty acid (VFA) end-product according to the following reaction:

$$C_6H_{12}O_6 + 4H_2O \rightarrow 2CH_3COO^- + 2HCO_3^- + 4H^+ + 4H_2 \quad (2.7)$$

The $\Delta G°'$ of this reaction is -206.3 kJ·mol^{-1} (Thauer et al. 1977) indicating that product formation is favorable at standard conditions (pH = 7). Production of butyrate also often occurs in H_2 producing organisms:

$$C_6H_{12}O_6 + 2H_2O \rightarrow CH_2CH_2CH_2COO^- + 2HCO_3^- + 3H^+ + 2H_2 \quad (2.8)$$

for this reaction $\Delta G^{\circ\prime} = -254.8\ kJ \cdot mol^{-1}$ (Thauer et al. 1977). Although the free energy change in butyrate formation is more favorable, it results in a less favorable H_2 yield than acetate formation; only 2 moles of H_2 are produced rather than 4 moles. The theoretical yield of 4 mol H_2 per mol glucose in eq 2.7 is not always attained in practice. In many organisms, both acetate and butyrate are produced and the experimental H_2 yield falls between the two theoretical values. When other end–products such as propionate, lactate, and solvents are produced, H_2 is not produced; therefore overall H_2 yields are decreased if these other co-products are present and cell growth also reduces H_2 yields.

It should be noted that yields greater than 4 mol H_2/mol glucose have been observed *in vitro* using free enzyme systems. Woodward et al. (2000) used 11 enzymes of the pentose phosphate pathway to convert glucose-6-phosphate to H_2 and Zhang et al. (2003) employed a 13 enzyme system to convert starch to H_2. However, these *in vitro* systems had extremely slow H_2 production rates and low H_2 partial pressures, which make them infeasible for economical H_2 production (Hallenbeck and Benemann 2002; Nath and Das 2004).

There are two specific mechanisms that organisms use for enzymatic conversion of pyruvate to H_2 (Hallenbeck 2005). The first mechanism is common in enteric bacteria, which are Gram-negative, rod-shaped, facultative anaerobes, and involves 2 sequential steps. Pyruvate is converted to formate and acetyl-CoA by the pyruvate:formate lyase (PFL) enzyme. Then formate:hydrogen lyase (FHL), which is not a single enzyme entity, but rather a formate dehydrogenase and hydrogenase that act together (Gottschalk 1986), splits formate to produce H_2 and CO_2. The reactions are as follows:

$$\text{Pyruvate} + \text{CoA} \xrightarrow[\text{PFL}]{} \text{Acetyl-CoA} + \text{Formate} \qquad (2.9)$$

$$\text{Formate} \xrightarrow[\text{FHL}]{} 2CO_2 + 2H_2 \qquad (2.10)$$

The second type of mechanism is found in strict anaerobes. It is referred to as the phosphoroclastic reaction, because phosphate is required for cleavage of pyruvate. First, in two steps pyruvate:ferredoxin oxidoreductase (PFOR) catabolizes pyruvate to acetyl-CoA and reduced ferredoxin (Fd) (eq 2.11 and 2.12) using thiamine pyrophosphate-containing oxidoreductase (TPP-E) and hydroxyethyl-TPP-E (HETPP-E) for phosphate transfer. Then hydrogenase (Hyd) accepts electrons from Fd and reduces protons to H_2 (eq 2.13). The final reaction involves phosphotransacetylase (PTA) which converts acetyl-CoA to acetyl phosphate (eq 2.14). The series of reactions are as follows:

$$\text{Pyruvate} + \text{TPP-E} \xrightarrow[\text{PFOR}]{} \text{HETPP-E} + CO_2 \qquad (2.11)$$

$$\text{HETPP-E} + \text{Fd(ox)} + \text{CoA} \xrightarrow[\text{PFOR}]{} \text{TPP-E} + \text{Acetyl-CoA} + \text{Fd(red)} \qquad (2.12)$$

$$\text{Fd(red)} + 2H^+ \xrightarrow[\text{Hyd}]{} H_2 + \text{Fd(ox)} \qquad (2.13)$$

$$\text{Acetyl-CoA} + P_i \xrightarrow[\text{PTA}]{} \text{Acetylphosphate} + \text{CoA} \qquad (2.14)$$

In some organisms flavodoxin is present as the electron carrier rather than ferredoxin (Gottschalk 1986). Since 2 pyruvate molecules are produced per molecule of glucose, 2 mol H_2 are produced from the phosphoroclastic reaction. An additional 2 mol H_2 can also be produced from glycolysis during substrate phosphorylation. The phosphorylation of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate yields NADH which can then donate electrons to form H_2. Two moles of H_2 can be produced from substrate phosphorylation, because two glyceraldehyde-3-phosphate moles are produced per mole of glucose consumed. More detail is provided on specific enzymes and pathways related to H_2 production in section 2.2 of this review.

2.1.1.2 Organisms

A variety of organisms are capable of H_2 production via fermentation pathways. Research in the field has not yet converged upon an optimal system, so pure cultures, mixed cultures, and defined co-cultures are all under study. Organisms with different O_2 and temperature requirements are being assessed. The primary metric for an optimal H_2 producer is the H_2 yield coefficient, which is generally defined as the moles of H_2 produced per mole of substrate consumed and represented by $Y_{H_2/Substrate}$. The model substrate that is used for comparison is glucose and more complex sugars are usually converted into glucose units for comparison between studies. Table 2.1 contains a non-exhaustive list of fermentation organisms and the yield coefficients from glucose, $Y_{H_2/Glucose}$,which have been adapted from other reviews (Hallenbeck 2005; Kapdan and Kargi 2006). For the sake of brevity only studies where the yield coefficients were greater than 2 mol H_2/mol glucose are included in Table 2.1, although H_2 producers with lower yields are still of interest, such as *Clostridium thermocellum* (Levin et al. 2006), which has the advantage of directly degrading lignocellulosic materials with its extracellular enzymes.

The H_2 production studies in Table 2.1 include both pure and mixed culture systems. Pure culture systems have the advantage of having one clearly defined species, which allows for a greater degree of process control and reduced concern that population dynamics may change drastically. Pure culture systems also facilitate the application of metabolic engineering tools to improve H_2 yields. Pathway manipulation is much simpler in one organism rather than a community. For example, Kumar et al (2001) demonstrated metabolic engineering for improved H_2 yields with a pure culture of *Enterobacter cloacae*, by using a substrate suicide technique to block alcohol formation pathways and a proton suicide technique to block organic acid

production. The increased yield is shown in Table 2.1. Disadvantages in using pure cultures are that a single culture may not be as robust to environmental perturbations as a mixed microbial community and it is important to maintain aseptic conditions to prevent contamination.

Advantages of using mixed cultures include the robustness of the system and the ease of inoculation. Seed cultures can be obtaining from sources such as sewage sludge and compost. The inoculum source is often heat-treated to enrich for H_2-producing spore-forming organisms, typically *Clostridia* sp., that will survive high temperature treatment and then become metabolically active again at lower growth temperatures (Hawkes et al. 2002; Angenent et al. 2004; Kapdan and Kargi 2006). Disadvantages to working with undefined mixed consortia for a fermentative H_2 system appear to outweigh the benefits. The primary disadvantage is that it can be difficult to control the population composition. For example, in mixed cultures it is important to prevent methanogens from becoming competitive with the H_2-producing species, because then methane production will become the dominant biological process rather than H_2 generation (Angenent et al. 2004). This means that process control is a major issue in these systems. Another issue is that several H_2-producing species that dominate in mixed consortia, e.g. *Clostridia* sp., produce solvents rather than VFAs during stationary growth phase (Hawkes et al. 2002). Methods to address the issues of controlling methanogenesis and solventogenesis include operation at low pH and high dilution rate (Chen et al. 2001; Hawkes et al. 2002; Oh et al. 2003; Angenent et al. 2004). Defined co-cultures may be a successful compromise between mixed and pure cultures, allowing for increased control and robustness of the system and metabolic engineering manipulation (further discussion is in section 2.1.1.5).

In selecting organisms for H_2 production, O_2 requirements are another issue to consider. Strict anaerobes such as *Clostridia* sp. require that no oxygen is present.

Table 2.1 Fermentative H_2 producing organisms

Organism2	$Y_{H_2/Glucose}$ (mol/mol)	Reference
Mesophiles		
Mixed Culture	2.1	Fang et al. (2002)
Sewage sludge seed	2.1	Fang et al. (2002)
Anaerobic sewage sludge	2.4	Lin and Lay (2004)
Food waste culture	2.5	Noike and Mizuno (2000)
Enterobacter cloacae IIT-BT 08	2.2	Kumar and Das (2000)
Enterobacter cloacae mutant	3.4	Kumar et al. (2001)
Ruminococcus albus	2.4	Iannotti et al. (1973)
Clostridium butyricum	2	Saint-Amans et al. (2001)
	2.2	Kataoka et al. (1997)
Citrobacter sp. Y19	2.5	Oh et al. (2003)
Clostridium sp.	2.4	Taguchi et al. (1995)
Clostridium acetobutyricum	2	Chin et al. (2003)
Co-culture: *C. butyricum* &	2.6	Yokoi et al. (1998)
Enterobacter aerogenes	2.7	Yokoi et al. (2002)
Thermophiles		
Thermotoga maritima	4	Schröder et al. (1994)
Pyrococcus furiosus	3	Kengen and Stams (1994)
		Schäfer and Schönheit (1992)
Spirochaeta thermophila	3	Janssen and Morgan (1992)
Thermotoga sp.	3.6	Janssen and Morgan (1992)
Spirochaeta sp.	2	Rainey et al.(1991)
Mixed culture	2.4	Ueno et al. (1995)
Thermoanaerobacterium saccharolyticum	2.4	Ueno et al. (2001)
Sludge compost	2.1	Morimoto et al. (2004)
Acetothermicus paucivorans	3.5	Dietrich et al. (1988)
Acetomicrobium flavidum	4	Soutschek et al. (1984)

Prevention of O_2 exposure could be difficult on an industrial scale and requires expensive reducing agents (Kapdan and Kargi 2006); however potential H_2 yields in strict anaerobes are considerably higher than yields in facultative anaerobes that are O_2 tolerant, such as *Enterobacter* sp. Microaerobic species, those that require O_2 partial pressures below atmospheric levels (Madigan et al. 2003), have also been described such as *Klebisalle oxytoca* HP1 (Minnan et al. 2005). Debate has arisen in the literature over the O_2 tolerance and potential for aerobic metabolism of *Thermotogales* sp. (Van Ooteghem et al. 2002; Eriksen et al. 2008; Le Fourn et al. 2008; Tosatto et al. 2008) and is discussed in greater detail in section 2.1.2.

A third issue to consider for organism selection is growth temperature. Both mesophilic and thermophilic organisms have been studied and are listed in Table 2.1. Mesophiles have optimal growth temperatures between 20 and 45 °C and thermophiles have optimal growth temperatures between 45 and 80 °C (Madigan et al. 2003). Hyperthermophiles are a subset of thermophilic organisms with an optimum temperature of 80 °C or greater (Kelly and Adams 1994; Amend and Shock 2001; Madigan et al. 2003).

There are several advantages for using thermophilic organisms for H_2 production systems. Elevated temperatures create a selection pressure for H_2 producers and therefore there is limited potential for contamination from unwanted species (Claassen et al. 1999). Hyperthermophiles also produce extracellular enzymes that can degrade complex biopolymers, such as starch, cellulose, hemicelluloses, etc, simplifying biomass degradation (Claassen et al. 1999). A third advantage of high temperature growth is that as temperature increases the affinity of the hydrogenase enzyme for H_2 decreases (Adams 1990; Claassen et al. 1999), so at higher temperatures H_2 is more likely to be evolved rather than consumed. Product formation also becomes more thermodynamically favorable at elevated temperatures,

because the change in Gibbs free energy, ΔG, of a reaction becomes more negative. This relationship is clear from the equation:

$$\Delta G = \Delta H - T\Delta S \quad (2.15)$$

where:

ΔG = change in Gibbs free energy of the reaction ($kJ \cdot mol^{-1}$),

ΔH = change in enthalpy of the reaction ($kJ \cdot mol^{-1}$),

T = temperature (K) of the reaction, and

ΔS = change in entropy from the reaction ($kJ \cdot mol^{-1} \cdot K^{-1}$).

The increased favorability of product formation with increased temperature was demonstrated by Lee and Zinder (1988), they determined that H_2 formation from acetate increased with increasing growth temperature in an acetate oxidizing methanogenic co-culture. The increased favorability of H_2 production at thermophilic temperatures is also apparent from the comparison of mesophilic and thermophilic yield coefficients (Hallenbeck 2005) , see Table 2.1.

2.1.1.3 Substrates

A variety of mixed and pure substrates have been used in H_2 production studies. Since the long-term objective is to apply fermentative H_2 systems to waste streams, many researchers have focused on the use of complex substrates such as waste materials including: dairy industry, olive mill, baker's yeast and brewery wastewaters (Kapdan and Kargi 2006), food processing wastewater (Oh and Logan 2005), bean curd manufacturing waste (Hallenbeck 2005), and paper sludge hydrosylate (Kadar et al. 2003; Kadar et al. 2004). Starch and lignocellulosic agricultural materials are also under investigation including rice bran, wheat bran

(Hallenbeck 2005), pretreated corn-stover (Datar et al. 2007), and *Miscanthus* (de Vrije et al. 2002; de Vrije et al. 2009).

Although mixed substrate studies will be important in the long-term, the use of well-defined pure substrates is necessary to develop mathematical models that can be used to manipulate metabolic pathways for H_2 production. Pure carbohydrate substrates under study include monomeric sugars such as glucose, fructose, xylose, and arabinose. More complex molecules such as polysaccharides, cellulose and starch, and disaccharides, specifically sucrose and maltose, have also been tested for H_2 yields (Hallenbeck 2005). An important consideration in substrate selection is that the substrate glucan linkage type may have an effect on the amount of H_2 produced. In a recent study on *Pyrococcus furiosus* Chou et al. (2007) determined that cellobiose grown cells had a higher yield (3.8 mol H_2/mol glucose equivalent) than maltose grown cells (2.6 mol H_2/mol glucose equivalent) although growth yields from the substrates were comparable.

2.1.1.4 Yield Improvement Strategies

The key issue to address for fermentative H_2 production is improving H_2 yields (Nath and Das 2004) and several strategies have been developed. Perhaps the simplest method to improve a fermentation process is to manipulate the environmental variables. In the case of H_2 production processes these would include simple variables such as temperature, pH, and substrate concentration, and the more complex task of optimizing mass transfer in the gas and liquid phases of the system. In H_2 production systems the gas phase partial pressure of H_2 is a limiting factor for H_2 production. Angenent (2004) derived the following equation from the Nernst equation to estimate

the maximum H_2 concentration allowable for an electron carrier to donate electrons to a hydrogenase and form H_2.

$$P_{H_2,max} \leq \exp\left\{\frac{2F\left(E^{\circ'}_{H_2} - E^{\circ'}_{x}\right)}{RT}\right\} \quad (2.16)$$

where:

$E^{\circ'}_{x}$ = redox potential of the electron carrier (mV),

$E^{\circ'}_{H_2}$ = redox potential of H_2 (mV),

F = Faraday's constant ($C \cdot mol^{-1}$),

R = ideal gas constant ($J \cdot K^{-1} \cdot mol^{-1}$),

T = temperature (K), and

$P_{H_2,max}$, the maximum partial pressure of H_2 (atm), is defined as:

$$P_{H_2,max} = \frac{K_{H_2}}{\text{fugacity coefficient}} \quad (2.17)$$

Fugacity is defined as the tendency of a compound to change from one state to another and the fugacity coefficient is fugacity/pressure, which for an ideal gas is assumed to be 1 atm^{-1}. In eq 2.17, K_{H_2} is the activity coefficient for the production of H_2. To obtain eq 2.16, equilibrium is assumed, meaning that there are equivalent concentrations of the reduced and oxidized forms of the electron carrier in the system, which may not be true *in vivo*. For the case of ferredoxin, $E^{\circ'}_{Fd} \approx$ - 400 mV depending on origin (Angenent et al. 2004), and given $E^{\circ'}_{H_2}$ = - 414 mV and standard conditions (T = 298 K), using eq 2.16 the acceptable $P_{H_2,max}$ is 0.3 atm ($3x10^4$ Pa). As mentioned earlier, with increased temperature, $P_{H_2,max}$ increases, however H_2 partial pressure is still a limiting factor even for hyperthermophiles. Methods to reduce the partial pressure include gas sparging, to actively displace H_2 from the gas and liquid phases, and reactor design, such as rhomboid and tapered reactors (Kumar and Das 2001) and trickling filter systems (van Groenestijn et al. 2002).

Although manipulation of environmental parameters can certainly improve H_2 yields, this sort of manipulation will not allow yields to exceed the Thauer theoretical limit of 4 mol H_2 / mol glucose (Thauer et al. 1977; Levin et al. 2004; Nath and Das 2004; Hallenbeck 2005). To improve H_2 yields beyond this limit metabolic engineering would be required, including both genetic changes to alter enzyme functions and the addition and/or deletion of entire metabolic pathways (Keasling et al. 1998). Despite the potential of using metabolic engineering for increased H_2 yields, no engineered organism has been experimentally shown to exceed this limit. The nascent field of systems biology provides a new strategy for taking metabolic engineering a step further. In systems biology mathematical modeling is used to describe not only the metabolic network of an organism, but also the interactions of metabolism with the transcriptional and regulatory networks. A system of linear equations can be developed to describe the network of reactions in an organism and the links between those reactions and the genome. Computational simulations can then be executed to predict the results of genetic and metabolic manipulations and the effect on the organism, offering a more comprehensive, guided, and inexpensive approach to experimental design than a traditional trial and error approach to genetic manipulation. More detail on the application of systems biology to improve hydrogen production in *Thermotogales* is in section 2.4.

In addition to metabolic engineering strategies to improve H_2 yields, multi-stage systems would be ideal to fully extract the maximum amount of H_2 available in a given carbohydrate feedstock. Two-stage fermentation systems have been proposed wherein the first stage would generate H_2 and CO_2 via dark fermentation of carbohydrate-based feedstocks and in the second stage photo-fermentation organisms would liberate the remaining H_2 and CO_2 from the organic acids generated in the first stage (Nath and Das 2004; Kapdan and Kargi 2006; Nath et al. 2008). Several groups

are also developing methods to couple H_2 fermentation systems to microbial fuel cells (Angenent et al. 2004; Logan and Grot 2006; Rupprecht et al. 2006). At first glance, it appears that multi-stage processes would most likely be spatially-separated, but it may also be possible to have multi-stage H_2 production occur in the same reactor in a co-culture system with paired organisms. Co-culturing has been investigated from several angles, combinations include fermentations of heterotrophic and photoheterotrophic organisms (Yokoi et al. 1998; Kawaguchi et al. 2001; Kapdan and Kargi 2006) and anaerobic and aerotolerant organisms, such as *Clostridium butyricum* and *Enterobacter aerogenes*, (Yokoi et al. 1998; Yokoi et al. 2002; Kapdan and Kargi 2006). Demain et al. (2005) discussed the potential for an ethanol producing co-culture using *Clostridium thermocellum*, an organism that can degrade cellulose and hemicelluloses, but can only consume hexoses, and another ethanol-producing species to consume the pentoses. Such a co-culture system could also be applied to H_2 production. *Clostridium thermocellum* is also a H_2 producer and could be coupled with *Thermotoga neapolitana* to convert the recalcitrant pentoses to H_2. Temperature could be a controlling factor in such a system, because *T. neapolitana* grows optimally at higher temperatures than *C. thermocellum*, which could potentially survive in an inactive state at these elevated temperatures due to its spore-forming capability.

2.1.2 *Thermotoga neapolitana* for H_2 Production

The thermophilic bacterial species *Thermotoga neapolitana* NS-E (DSM 4359) was first isolated in 1986 in a submarine hot spring in the Bay of Naples, Italy (Belkin et al. 1986), but was not officially named until 1988 when it was determined to be a distinct species from *Thermotoga maritima* MSB8 (DSM 3109) (Jannasch et al. 1988). The genus name *Thermotogales* and the species *T. maritima* had also been initially

proposed in 1986 after the isolation of the type strain of this species from geothermally heated sea sediments from Italy, Vulcano, Porto di Levanti (Huber et al. 1986). *Thermotogales* was selected as the genus name for these organisms based on their growth at high temperatures (thermo-) and the sheath-like structure surrounding the rod-shaped cells, which resembles a toga. Jannasch et al. (1988) distinguished between the two species with the following basis: compared to *T. maritima*, *T. neapolitana* was immotile, had increased growth rates at 85 °C, a lower G + C content, and there was only 25% homology in DNA-DNA hybridization between the two species.

The temperature range for *T. neapolitana* growth is 55 – 90 °C (Jannasch et al. 1988). The optimal growth temperature, which is defined as the temperature when doubling time is minimal, was determined to be around 80 °C by Jannasch et al. (1988) and 77 °C by Childers et al. (1992). A pH of approximately 7 - 7.5 is reported to be optimal for growth (Belkin et al. 1986; Jannasch et al. 1988). A variety of carbohydrate substrates are consumed by *T. neapolitana* including ribose, xylose, glucose, sucrose, maltose, lactose, galactose, starch, and glycogen (Jannasch et al. 1988). Childers et al. (1992) found that starch resulted in higher cell densities than growth on lactose or glucose. Van Ooteghem (2004) used Biolog AN microplates to test 95 different substrates selected by the manufacturer (Biolog). Some of the notable results are shown here in Table 2.2. The 100% positive results indicate that all replicated plates in the study, n = 20 for H_2 and n= 40 for N_2, showed respiration. *T. neapolitana* utilized all tested peptides except for glycyl-L-proline. It is interesting that L-alanine is utilized as a substrate, because it has also been reported to be a metabolic product of *T. neapolitana* and other members of the *Thermotogales* (Ravot et al. 1996) as well as the archaeal species *Pyrococcus furiosus* (Kengen and Stams 1994). Other *T. neapolitana* metabolic products, acetic acid and lactic acid, were not

Table 2.2 Notable substrate utilization results for *T. neapolitana* as determined by Van Ooteghem et al. 2004. Biolog plates testing for respiration on 95 different substrates were incubated under either N_2 or H_2 atmospheres.

100% Positive results under N_2 and H_2	**Utilized peptides**	**Non-utilized substrates**
D-Fructose	L-Alanine	Succinamic acid
L-Fucose	L-Alanyl-L-glutamine	Glycyl-L-proline
L-Galactose	L-Alanyl-L-glutamine	Adonitol
D-Mannose	L-Asparagine	Acetic acid
L-Rhamnose	L-Glutamic acid	Formic acid
Pyruvate	L-Glutamine	α-Hydroxybutyric acid
	Glycyl-L-aspartic acid	β-Hydroxybutyric acid
	Glycyl-L-glutamine	DL-Lactic acid
	Glycyl-L-methionine	L-Lactic acid
	L-Methionine	D-Saccharic acid
	L-Phenylalanine	Succinic acid
	L-Serine	*meso*-Tartaric acid
	L-Threonine	DL-Lactic acid methyl ester
	L-Valine	Urocanic acid
	L-Valine + L-aspartic acid	

degraded in any of the assays.

Limited analysis of degradation of complex biomass substrates has been done with *Thermotoga* species. Growth on pretreated *Miscanthus* has been tested with *Thermotoga elfii* (de Vrije et al. 2002) and *T. neapolitana* (de Vrije et al. 2009). In the work with *T. neapolitana* glucose and xylose mixtures in a 7:3 (glucose:xylose) ratio were tested and *T. neapolitana* simultaneously degraded glucose and xylose (de Vrije et al. 2009). We have also seen simultaneous degradation in batch cultures with 1:1 ratios (unpublished work). In the analysis of growth on *Miscanthus* hydrosylate, it was found that *T. neapolitana* consumes furfural and 5-hydroxymethyl furfural (HMF), toxic byproducts of acidic pretreatment of lignocellulosic material, and can tolerate up to 4 g/L of these compounds (de Vrije et al. 2009).

The fermentation products that have been identified for *T. neapolitana* include: H_2, CO_2, acetate, lactate, alanine, and, if inorganic sulfur is provided asthe sulfur source, H_2S. Although some of these products have been analyzed, for example Van Ooteghem et al. (2002) analyzed H_2 and CO_2 gas concentrations in the headspace of serum bottle cultures and Ravot et al. (1996) analyzed acetate, alanine, H_2, and H_2S when thiosulfate was included in the medium, a closed carbon balance on this species is not yet available in the literature, but is reported here in Chapter 3.

One significant difference between the metabolism of *T. neapolitana* and *T. maritima* is that *T. neapolitana* can grow in a defined medium with cysteine as a sulfur source (Childers et al. 1992), but *T. maritima* cannot (Schröder et al. 1994). *T. maritima* requires either a complex peptide source such as yeast extract or elemental sulfur. The use of elemental sulfur results in the production of H_2S instead of H_2 – not a desirable trait for H_2 production. The inclusion of an undefined medium component such as yeast extract can be problematic for obtaining consistent results in

physiological studies and also cannot be accounted for in metabolic models (Palsson 2006).

The oxygen sensitivity of *T. neapolitana* has become a topic of debate in the literature. Belkin et al. (1986) reported that cells were insensitive to O_2 at temperatures below the permissive growth range and that cells retained viability after storage in air for several months at 4 °C. However, Belkin et al. also observed that at permissive growth temperatures cell death occurred rapidly in the presence of oxygen (1986). Later work by Childers et al. (1992) confirmed these observations made by Belkin et al. (1986). Childers et al. (1992) developed a defined marine salts medium for *T. neapolitana* liquid cultures and also were able to plate the organism at room temperature on the bench without anoxic conditions. However, anaerobic conditions were required for incubation at growth temperatures. Van Ooteghem et al. (2002) cultured *T. neapolitana* under microaerobic conditions (~6% O_2 headspace concentration) and observed decreased O_2 concentrations over time as H_2 and CO_2 concentrations increased in growing cultures. This led to the hypothesis that *T. neapolitana* was not only O_2 tolerant, but also performing aerobic metabolism with increased H_2 production. Van Ooteghem et al. (2002) calculated hydrogen yield coefficients that were much higher than the expected 4 mol H_2/ mol glucose. Based on this research (Van Ooteghem et al. 2002) a patent was obtained on behalf of the U.S. DOE, with Van Ooteghem as inventor, for H_2 production under microaerobic conditions using *Thermotogales* sp. (Van Ooteghem 2005). Unfortunately the elevated H_2 yields were probably due to inaccurate glucose measurements, because my results (Munro et al. 2009) and the results of Eriksen et al. (2008) negate the results of Van Ooteghem et al. (2002). Eriksen et al. (2008) showed that the cysteine present in the growth medium was capable of reducing O_2 which would explain the decrease in O_2 concentration that Van Ooteghem et al. observed. Our results and the

Eriksen et al. study (2008) also determined that if O_2 is injected into actively growing anaerobic cultures, H_2 yields were lower than under anaerobic conditions, not elevated as Van Ooteghem indicated (2002). It can be concluded from these results that *T. neapolitana* is not a microaerobic species.

Concomitantly with publication of the Eriksen study (2008), Tosatto et al. (2008) published a study comparing the structural characteristics of a *T. maritima* hydrogenase and a hypothetical *T. neapolitana* hydrogenase to identify characteristics that would render the *T. neapolitana* hydrogenase less sensitive to O_2 than the *T. maritima* hydrogenase. No claims were made regarding microaerobic metabolism in this study, instead it was suggested that amino acid substitutions in a model of *T. neapolitana* hydrogenase compared with the solved structure of *T. maritima* hydrogenase were the reason for the decreased O_2 sensitivity in *T. neapolitana* when compared with *T. maritima*. It should be noted that the differences in O_2 sensitivity between the two organisms have not been directly compared or published. Tosatto et al. (2008) based their theoretical study on personal communication with Van Ooteghem regarding the O_2 sensitivity of the two organisms. However, it is possible that both Eriksen et al. (2008) and Tosatto et al. (2008) are correct and *T. neapolitana* does not metabolize O_2, but its hydrogenase has decreased sensitivity to O_2 compared to hydrogenases in related species.

In another recent study, Le Fourn et al. (2008) examined the O_2 tolerance of *T. maritima* and found that with up to 0.5% (v/v) O_2 exposure growth was possible. Transcriptomic and proteomic results indicated that with a low O_2 concentration cells upregulated production of a flavoprotein that was found to have oxygen reductase activity. If O_2 tolerance is truly a trait of either species, *T. maritima* or *T. neapolitana*, the findings of Le Fourn et al. (2008) in *T. maritima* suggesting that tolerance of small amounts of O_2 is due to an oxygen reductase appears to be a more realistic explanation

than the Tosatto et al. (2008) hypothesis of hydrogenase-based resistance, and would support the O_2 tolerance observed in both *T. maritima* and *T. neapolitana*. This is primarily because Tosatto et al. (2008) base their hypothesis on the assumption that *T. maritima* lacks O_2 tolerance and a theoretical inference based on differences in hydrogenase structure between the two species, whereas Le Fourn et al. (2008) have biochemical results supporting scavenging of O_2 by an oxygen reductase in *T. maritima*. If an oxygen reductase was also observed in *T. neapolitana* then O_2 tolerance in both species could be attributed to this enzyme.

Although some basic questions regarding *T. neapolitana* metabolism have been answered and investigations have been made regarding H_2 yields and production rates in this organism (Van Ooteghem et al. 2002; Eriksen et al. 2008; Nguyen et al. 2008; Munro et al. 2009), much still remains to be discovered in *T. neapolitana*. If *T. neapolitana* is to become a viable organism for H_2 fermentation systems more research must be performed to optimize it. The following sections of this review will discuss the use of biochemical and functional genomics data of *Thermotogales* sp. to develop an *in silico* model representing *T. neapolitana* that can be used to optimize H_2 production in this organism.

2.2 *Thermotoga neapolitana* Carbohydrate Metabolism

For improved H_2 production in *T. neapolitana*, knowledge of the enzymes and pathways of central metabolism is required. There is biochemical and genetic information available that is specific to *T. neapolitana*; however, the majority of research on *Thermotoga* species has focused on *Thermotoga maritima* MSB8, a close relative of *T. neapolitana*. In this section (2.2), an overview is provided of the

available biochemical information in the literature pertaining to *T. neapolitana* carbohydrate metabolism, and is augmented by *T. maritima* results.

T. maritima has become the model system for investigating *Thermotoga* species largely due to the fact that the annotated genome sequence of *T. maritima* was published first (Nelson et al. 1999). However, the draft genome annotation for *T. neapolitana* has recently become publicly available (Lim et al. 2009) and can be used for comparative analysis to the more well-studied *T. maritima*. Fermentation metabolite studies of both species (Schröder et al. 1994; Ravot et al. 1996; Munro et al. 2009) indicate that, at least with respect to central metabolism, these organisms are very similar and inferences can be made regarding *T. neapolitana* metabolism based on *T. maritima* metabolism. In addition, genome alignment work both by DeBoy et al. (2006) and described here in Chapter 4, indicate a high degree of similarity between the two genomes. DeBoy et al. (2006) performed a whole genome alignment between *T. maritima* MSB8 and *T. neapolitana* NS-E and found that 1,726 proteins are shared between the two strains, out of 1,838 total proteins for *T. maritima* MSB8 and 1,903 proteins for *T. neapolitana* NS-E. There are large scale inversion events in the genomes, but despite this rearrangement, the two proteomes have an average percentage of identity of 83.6% and an average percentage of similarity of 92.4%. The Mauve Multiple Genome Alignment software package (Darling et al. 2004) was used to align the two genomes and develop a constraint-based model for *T. neapolitana*. This work is described in Chapter 4 of this dissertation.

Competence, the ability of a cell to take up DNA from the extracellular environment, has not been demonstrated in *T. maritima* or *T. neapolitana* although genes encoding type II secretion pathway proteins and type IV pili were identified in the *T. maritima* genome annotation (Nelson et al. 1999; Conners et al. 2006). Effort has gone into developing a genetic system for *Thermotoga* species, but a stable system

has not been created (Noll and Vargas 1997; Conners et al. 2006). Childers et al. (1992) developed a defined minimal medium and were able to cultivate *T. neapolitana* on solid media. In 1994, they developed antimetabolite resistant and auxotrophic mutants (Vargas and Noll 1994). Yu et al. (2001) made the most progress by creating a liposome-mediated transfer system with a plasmid isolated from the strain *Thermotoga* sp. RQ2. Unfortunately, although initial transfer of DNA using this procedure was successful, the DNA transfer was transient - antibiotic resistance conferred by the plasmid was lost after 20 generations. Due to the lack of a robust genetic system for *Thermotoga* species, the use of *Escherichia coli* host strain libraries had been the primary tactic for studying genes and enzymes in these species. However, with the availability of the *T. maritima* MSB8 genome annotation, functional genomics studies have emerged as the leading methodology to explore *T. maritima* MSB8 and update the genome annotation. Phylogenetic studies have been performed using *T. maritima* MSB8 as a standard for comparison to other *Thermotoga* strains including *T. neapolitana* NS-E (DSM 4539) (Nesbø et al. 2001; Nesbø et al. 2002; Mongodin et al. 2005; Nesbø et al. 2006). Many of these studies have a focus on topics that are related to evolution, such as lateral gene transfer between members of *Thermotogales* and between *Thermotogales* and archaeal species (Nelson et al. 1999; Nesbø et al. 2001; Conners et al. 2006). Some of the papers explore the definition of a species (Nesbø et al. 2006). These phylogenetic and evolutionary studies utilize suppressive subtractive hybridization (SSH) (Nesbø et al. 2002) and comparative genomic hybridization (CGH) (Mongodin et al. 2005) to probe the similarities and differences between *Thermotoga* genomes.

Functional genomics studies have also been used to answer more specific questions regarding *T. maritima*, such as elucidating regulation of carbohydrate metabolism and transport (Chhabra et al. 2003; Nguyen et al. 2004; Conners et al.

2005; Nanavati et al. 2005), heat-shock response (Pysz et al. 2004), and behavior in communities (Pysz et al. 2004). These studies tend to specifically focus on *T. maritima* with the goal of improving the annotation of its genome, but the information may also be applicable to *T. neapolitana*.

The impetus for functional genomics efforts in *T. maritima* came from difficulties in identifying regulatory mechanisms in *T. neapolitana*. Catabolite repression studies with *T. neapolitana* showed that growth on lactose is suppressed by the presence of glucose and although the regulation system was not identified, it was found to be independent of cAMP (Vargas and Noll 1996). Related studies on glucose and galactose transport were conducted (Galperin et al. 1996; Galperin et al. 1997), and it was determined that regulation of substrate utilization was controlled at the transcriptional level. In a follow-up study, Nguyen et al. (2001) quantified *T. neapolitana* total RNA of bgalA, bgalB, bglA, and bglB, after growth on different carbon substrates, but they were unable to isolate specific regulatory mechanisms. They concluded that a broader spectrum of genes needed to be investigated, thus they proposed switching to *T. maritima* as an experimental organism, because of its available genome annotation. Several functional genomics studies from the Kenneth Noll and Robert Kelly research groups have since been published focusing on carbohydrate metabolism and transport with the primary results of establishing the identity of genes and the related products (Chhabra et al. 2003; Nguyen et al. 2004; Conners et al. 2005; Nanavati et al. 2005).

2.2.1 Extracellular Catalysis of Polysaccharides

T. neapolitana is capable of degrading α-linked and β-linked polysaccharides, and the resulting monomer and oligomer hydrolysis products. Prior to transport across

the cell membrane larger polysaccharides require extracellular catalysis. Further catalysis of oligosaccharides occurs within the cell and is discussed in section 2.2.3.

2.2.1.1 α-Linked Polysaccharides

T. neapolitana uses starch, a polysaccharide with mixed α-1,4- and α-1,6-linkages, as a growth substrate (Jannasch et al. 1988) and even shows improved growth on starch, when compared with growth on monosaccharides (Childers et al. 1992). However, extracellular amylases have not been isolated in *T. neapolitana*, although intracellular enzymes that process the maltodextrin products of starch degradation have been purified from this organism (Berezina et al. 2003; Lunina et al. 2003; Veith et al. 2003). Amylases have been found in *T. maritima*, however these amylases are not free extracellular enzymes, but remain bound to the toga outer membrane structure (Schumann et al. 1991). Liebl et al. (1997) examined the gene sequence and properties of *T. maritima* α-amylase (AmyA) and classified it as a putative lipoprotein that is localized to the exterior membrane. Although it is difficult to obtain significant quantities of the enzyme (Schumann et al. 1991; Liebl et al. 1997), when it was obtained the amylase activity was extremely high, specific activities of the purified enzyme with soluble starch and amylose were about 3,800 and 5,600 U/mg respectively (Liebl et al. 1997). Liebl et al. (1997) suggest that the scarcity of AmyA is related to its high activity; the high efficiency of the enzyme may mean that only a small quantity needs to be produced. Perhaps this scarcity is the reason such an enzyme has not been isolated from *T. neapolitana*. In addition to amylases, pullulanases have also been identified in *T. maritima* which catalyze the degradation of pullulan, a mixed α-1,4- and α-1,6-linked polymer of glucose (Bibel et al. 1998; Kriegshauser and Liebl 2000; Conners et al. 2006). Since amylases and

pullulanases are present in *T. maritima* it is likely that they are also produced by *T. neapolitana,* especially amylases considering *T. neapolitana* starch metabolism.

2.2.1.2 β-Linked Polysaccharides

T. neapolitana has extracellular cleavage activity for a variety of β-linked polysaccharides, including β-1,4-glucans, β-1,3-glucans and β-1,6-glucans, mannans, and xylan. For degradation of cellulose, which contains β-1,4-linked glucose subunits, *T. neapolitana* has two endocellulases, CelA and CelB (Bok et al. 1998; Liebl 2001). These endocellulases randomly cleave β-1,4-linkages along cellulose chains, as opposed to exocellulases (cellobiohydrolases), which only cleave cellobiose from the ends of β-1,4-linked chains. Bok et al. (1998) determined that CelA and CelB of *T. neapolitana* NS-E have 86% and 94% similarity to Cel12A and Cel12B of *T. maritima* MSB8. According to the CAZy glycoside hydrolase (GH) family characterization, CelA and CelB from *T. neapolitana* are also in the GH12 family. Expression of *T. neapolitana* CelA and CelB is repressed in the presence of glucose and β-glucosidase synthesis is induced by growth on cellobiose (Bok et al. 1998). In their study Bok et al. (1998) also found that the mRNA transcripts of CelA and CelB were monocistronic (two separate transcripts) rather than polycistronic as they had hypothesized. The specific regulation mechanism of CelA and CelB is still unclear although repression is cAMP independent (Vargas and Noll 1996).

In cellulose-degrading organisms, endocellulases act with exocellulases in a synergistic manner to degrade cellulose. Organisms that have a diverse array of exocellulases and endocellulases are often capable of degrading not just amorphous cellulose, but also the more recalcitrant crystalline cellulose. Extracellular cellobiohydrolase (exocellulase) activity has not been found in any *Thermotoga*

species which explains decreased activity of these species on crystalline cellulose (Conners et al. 2006, Yernool et al. 2000). Instead small oligomers and monomers (e.g. cellotetraose, cellotriose, cellobiose, glucose) generated by endocellulase activity are transported into the cell and intracellular processing enzymes further degrade the oligosaccharides to D-glucose (see sections 2.2.2 and 2.2.3 for detail).

β-1,3-glucan degradation has also been examined in *T. neapolitana* (Zverlov et al. 1997; Zverlov et al. 1997). Laminarin, a β-1,3-linked glucose polymer, is degraded by *T. neapolitana* laminarinase to laminaribiose, a dissaccharide. Zverlov et al. (1997) purified an endo-1,3-β-glucanase (laminarinase), LamA, and the protein BglB, which has both β-glucosidase and laminaribiase activity. Laminaribiase activity means that when BglB is coincubated with LamA, laminarin is completely hydrolyzed to glucose. Specifically, LamA releases laminaribiose disaccharides from laminarin and BglB then cleaves laminaribiose into glucose monomers, *in vivo* BglB activity is intracellular, so transport of laminaribiose is required prior to hydrolysis.

T. neapolitana is also capable of degrading polysaccharide components of hemicellulose, including mannan and xylan. Duffaud et al. (1997) first isolated from *T. neapolitana* the mannan degrading enzymes β-mannanase (1,4-β-D-mannan mannanohydrolase [EC 3.2.1.78]) and β-mannosidase (β-D-mannopyranoside hydrolase [EC 3.2.1.25]), as well as α-galactosidase (α-D-galactoside galactohydrolase [EC 3.2.1.22] which cleaves galactomannan sidechains. Parker et al. (2001) later determined that of these three enzymes only β-mannanase (the gene product of *man5*) has extracellular activity. β-mannanase, also called mannan endo-1,4-β-mannosidase, acts randomly along mannan, galactomannan, and glucomannan polymers. β-mannosidase and α-galactosidase are involved in intracellular processing of mannan (Parker et al. 2001).

Sunna et al. (1997) examined xylan degradation activity of *T. neapolitana*. Xylanases, β-xylosidases, and α-arabinofuranosidases were present in activity assays. No enzymes were purified in this study (Sunna et al. 1997), however two xylanases, XynA and XynB, have been isolated from *T. neapolitana* and have endoxylanase activity (Zverlov et al. 1996; Velikodvorskaya et al. 1997). Based on knowledge of xylanase structure in *T. maritima*, Mamo et al. (2007) attempted to increase catalysis of insoluble xylan by fusing an N-terminal CBM domain from *T. neapolitana* XynA, which only binds xylan, not cellulose, to an alkaline active xylanase from *B. halodurans* S7. The result was increased activity on insoluble xylan, although activity on soluble xylan did not improve over the control enzyme. In a recent study Liebl et al. (2008) showed that xylanases are associated with the toga structure in *T. maritima*. As mentioned previously, membrane-bound amylase activities have also been identified for *T. maritima.* Cellulosomes in mesophilic anaerobes are cell membrane associated rather than free in the extracellular environment and Gram-negative bacteria often have enzyme activity at the extracellular surface, therefore it is not surprising that the toga structure in *Thermotoga* species would contain enzymes with glycosyl hydrolase activity. The membrane association of these of enzymes is also interesting given that a third mechanism of cellulose degradation has been proposed in addition to free cellulase systems (with exocellulolytic components) and cellulosome activity. It has been suggested, based on the starch degradation mechanism in *Bacteroides thetaiotaomicron*, that this mechanism may involve a complex at the extracellular surface that binds cellulose and allows for transport into the periplasmic space where further degradation occurs (Wilson 2008). Organisms suspected to have this third mechanism, include *Cytophaga hutchinsonii* and *Fibrobacter succinogenes* (Wilson 2008), however growth on cellulose for these organisms is significantly faster than for *T. neapolitana*, which is evidence against

T. neapolitana using this as a cellulose-degrading mechanism although it may be the mechanism for starch degradation in *T. neapolitana*.

2.2.2 Transport

Transport proteins of *Thermotoga* species are an area under active study. A phosphotransferase system (PTS) is not present in *T. neapolitana* or *T. maritima* (Galperin et al. 1996; Nelson et al. 1999; Conners et al. 2006). In lieu of a phosphotransferase system, ABC (ATP-binding cassette) transporters are present and seem to be important for carbohydrate uptake (Nelson et al. 1999; Nanavati et al. 2002; Nanavati et al. 2005; Conners et al. 2006; Nanavati et al. 2006). ABC transporters consist of a periplasmic substrate binding protein (SBP), a membrane-spanning protein (MSP) that acts as a channel for transport, and a cytoplasmic ATP-binding protein (ABP) that hydrolyzes ATP to provide energy for transport. Many of the *T. maritima* ABC transport proteins and encoding genes have been identified. Nanavati et al. (2002; 2005) identified two sets of sugar family ABC-binding proteins, *malE2F2* for maltose, maltotriose and trehalose, and *malE1F1G1* for maltose, maltotriose, and β-1,4-mannotetraose. In a subsequent study Nanavati et al. (2006) determined the substrate specificity of more substrate binding proteins from the sugar ABC family of transport proteins and also determined that Opp/Dpp (oligopeptide/dipeptide) ABC family transporters, which appear to have been attained by horizontal gene transfer from archaea, bind carbohydrates including, laminarin, cellobiose, xylan, xylose, glucose oligosaccharides, mannobiose, and β-mannans. Conners et al. (2005) corroborated results from Nanavati et al. (2006) and added some additional transporters based on transcriptional data from microarray experiments with

14 different carbohydrate substrates, although these designations still require biochemical confirmation for permanence.

Although most of the biochemical and genomic evidence for transport proteins has been found in *T. maritima* through functional genomics studies, mechanistic studies of sugar transport by Noll and coworkers have focused on *T. neapolitana*. Galperin et al. (1996) used 2-DOG (a non-metabolizable glucose analog) and found that it was accumulated against a concentration gradient, which is indicative of glucose active transport. They confirmed the ATP-dependence of this transport, because activity was abolished by addition of sodium arsenate (an ATP-inhibitor) and pyruvate was required for transport. They also determined that the process was Na^+ dependent and there was no PTS present in this species, because phosphoenolpyruvate (PEP)-dependent PTS activity was not found in sugar phosphorylation assays, ATP was required instead. Also in 1996 Vargas and Noll conducted a study of substrate preferences in *T. neapolitana*. They determined that when presented with glucose and lactose, *T. neapolitana* exhibited diauxic growth. Lactose, galactose and cellobiose induced β-galactosidase activity and glucose repressed this activity only when induced with lactose or galactose. Vargas and Noll (1996) also measured cAMP levels in these experiments and found that it was absent; this indicates that the regulatory mechanism is cAMP-independent. They hypothesized that either inducer exclusion or expulsion are involved in catabolite repression. In a follow-up study on β-galactoside transport Galperin et al. (1997) ruled out both of these mechanisms. Glucose addition actually increased uptake of the nonmetabolizable β-galactoside analog methyl-β-D-thiogalactopyranoside (TMG), rather than decreasing the intracellular TMG concentration, which rules out inducer expulsion as a mechanism. Inducer exclusion was also absent, because glucose did not inhibit initial TMG uptake. β-galactoside transport showed ATP and Na^+ dependence and no PTS, just as in glucose transport,

however, contrary to the glucose transport system which was constitutive, Galperin et al. (1997) found that the β-galactoside transport system was inducible by lactose and galactose and repressed, but not inhibited by glucose. These results indicate that regulation of β-galactoside transport occurs at the genetic level, and does not involve regulation of the transport proteins themselves by inducer exclusion or expulsion.

2.2.3 Intracellular Processing of Oligosaccharides

2.2.3.1 α-Linked Oligosaccharides

Two α-glucosidases have been isolated in *T. neapolitana* in a maltodextrin utilization gene cluster, AglA and AglB (Berezina et al. 2003; Lunina et al. 2003). First, *aglA* and *aglB* were identified and sequenced in genomic DNA from *T. neapolitana* (Berezina et al. 2003) then *aglA* and *aglB* were expressed in *E. coli* and the properties of the recombinant enzymes were studied (Lunina et al. 2003). AglA and AglB have complementary substrate specificities on linear maltooligosaccharides, cyclic maltooligosaccharides, and di- and trisaccharides (Lunina et al. 2003). AglA, is a family 4 glycoside hydrolase with broad substrate specificity. AglA is active on di- and trisaccharides with different types of α-linkages including α-1,1-, α-1,2-, α-1,3-, α-1,4-, and α-1,6- linkages. AglB, is a family 13 glycoside hydrolase and has high activity on linear (maltotriose – maltoheptaose) and cyclic maltooligosaccharides. Although AglB can degrade linear malto-oligosaccharides, it lacks substrate specificity for maltose; however, AglA can cleave maltose. Veith et al. (2003) speculate that the complementary specificity of these enzymes is the reason why both are produced in *Thermotoga* species. In addition to intracellular α-glucosidase activity, *T. neapolitana* also has a 4-α-glucanotransferase (EC 2.4.1.25) that transfers a

segment of an α-1,4-D-glucan to a new position in an acceptor, which may be glucose or an α-1,4-D-glucan (Berezina et al. 1999).

2.2.3.1 β-Linked Oligosaccharides

T. neapolitana has several intracellular enzymes that are involved in the breakdown of β-linked oligosaccharides. As mentioned in section 2.2.1, *T. neapolitana* produces two endocellulases, CelA and CelB. It is clear that CelB has extracellular activity, because of its N-terminal localization sequence CelB is located either on or embedded in the outer membrane (toga), however the location of CelA in *T. neapolitana* is still unconfirmed. Evidence supporting designation of CelA as an intracellular enzyme include the lack of an N-terminal localization sequence (Yernool et al. 2000) and identification of the *T. maritima* Cel12A as an intracellular endocellulase by Connors et al. (2006).

Two key intracellular enzymes for processing of oligosaccharides were purified from *T. neapolitana* and characterized by Yernool et al. (2000), a 1,4-β-D-glucan glucohydrolase and a cellobiose phosphorylase. The 1,4-β-D-glucan glucohydrolase, GghA, cleaves glucose molecules from the ends of small 1,4-β-glucan oligomers. Using a Michaelis-Menten type model the catalytic efficiencies for different oligosaccharides showed the order of preference for GghA activity is cellotetraose > cellotriose > cellobiose. Catalytic efficiency is defined as k_{cat}/K_m where k_{cat} is the turnover number, the number of times an enzyme converts a unit of substrate to a unit of product per unit time and K_m is the Michaelis-Menten constant, the substrate concentration required to achieve half of the maximum reaction velocity. Cleavage of cellobiose by GghA is very slow; this low affinity for cellobiose is the reason that Yernool et al. (2000) designated it GghA rather than BglA (β-glucosidase)

although the two names are used interchangeably (McCarthy et al. 2004; Conners et al. 2006). BglA may become a more appropriate designation, because directed evolution of GghA has resulted in a 31% increase in catalytic efficiency (k_{cat}/K_m) of the enzyme on cellobiose compared with the wild-type enzyme (McCarthy et al. 2004).

The *T. neapolitana* cellobiose phosphorylase (CbpA) releases one glucose-1-phosphate and one glucose molecule per dissacharide. In this reaction the energy of the β-1,4-glycosidic bond is conserved, only one ATP molecule is needed to convert the glucose molecule to glucose-1-phosphate. Phosphoglucomutase converts glucose-1-phosphate to glucose-6-phosphate for glycolysis without consuming ATP. The presence of CbpA could be a form of ATP conservation in *T. neapolitana* (Yernool et al. 2000; Conners et al. 2006).

Based on the enzymatic activities of GghA and CbpA, Yernool et al. (2000) proposed a mechanism of cellulose degradation that compensates for the lack of cellobiohydrolase activity in *T. neapolitana*. Extracellular degradation may involve both CelB and LamA. CelB acts as an endoglucanase on cellulose β-1,4-glucan linkages and the laminarinase LamA acts on both β-1,4- and β-1,3-linkages. Following transport into the cell CelA acts as an endoglucanse to further reduce the length of the oligosaccharides, probably with assistance from BglB, which has both laminaribiase and 1,3-β-glucosidase activity. GghA (BglA) can then cleave glucose from the small oligosaccharides and CbpA generates glucose and glucose-1-phosphate from each molecule of cellobiose that is not cleaved by GghA.

Intracellular enzymes have also been identified in *T. neapolitana* for the degradation of oligosaccharides derived from hemicelluloses. For intracellular processing of mannan-derived oligosaccharides *T. neapolitana* uses a β-mannosidase and an α-galactosidase (Duffaud et al. 1997; King et al. 1998; Parker et al. 2001).

Intracellular enzymes related to xylan degradation in *T. neapolitana* have not been isolated by biochemical means. However, Sunna et al. (1997) found β-xylosidase activity in extracts of both *T. neapolitana* and *T. maritima* and a hypothetical annotation based on sequence analysis of the gene for an intracellular β-xylosidase is in the EMBL/GenBank/DDBJ databases (Yernool et al. 2002).

2.2.4 Monosaccharide Catabolism

2.2.4.1 Glycolysis

Work with *T. maritima* indicates that there are two possible forms of glycolysis that might be used by *T. neapolitana*, the Embden-Meyerhof-Parnas (EMP) pathway and the Entner-Doudoroff (ED) pathway. The EMP pathway, shown in Figure 2.1, is the more common pathway and is energetically more favorable (2 ATP are produced by the EMP pathway whereas only 1 ATP is produced from the ED pathway). Stage 1 of the EMP pathway is preparatory, involving phosphorylation of glucose with 2 ATP molecules and reduction from a 6-carbon molecule to a pair of 3-carbon intermediates, dihydroxyacetone-phosphate and glyceraldehyde-3-phosphate. Triosephosphate isomerase rapidly interconverts these 3-carbon intermediates, depending on cellular requirements for dihydroxyacetone-phosphate. It is the glyceraldehyde-3-phosphate form that is a substrate for stage 2 of the EMP pathway. Stage 2 is repeated twice for each glyceraldehyde-3-phosphate molecule generated from glucose to produce 2 pyruvate molecules. During stage 2 substrate-level phosphorylation occurs, in two separate reactions 3-phosphoglycerate kinase and pyruvate kinase each cleave a phosphate group to produce ATP. The total net ATP yield from EMP glycolysis is 2 ATP (4ATP are produced and 2ATP are consumed).

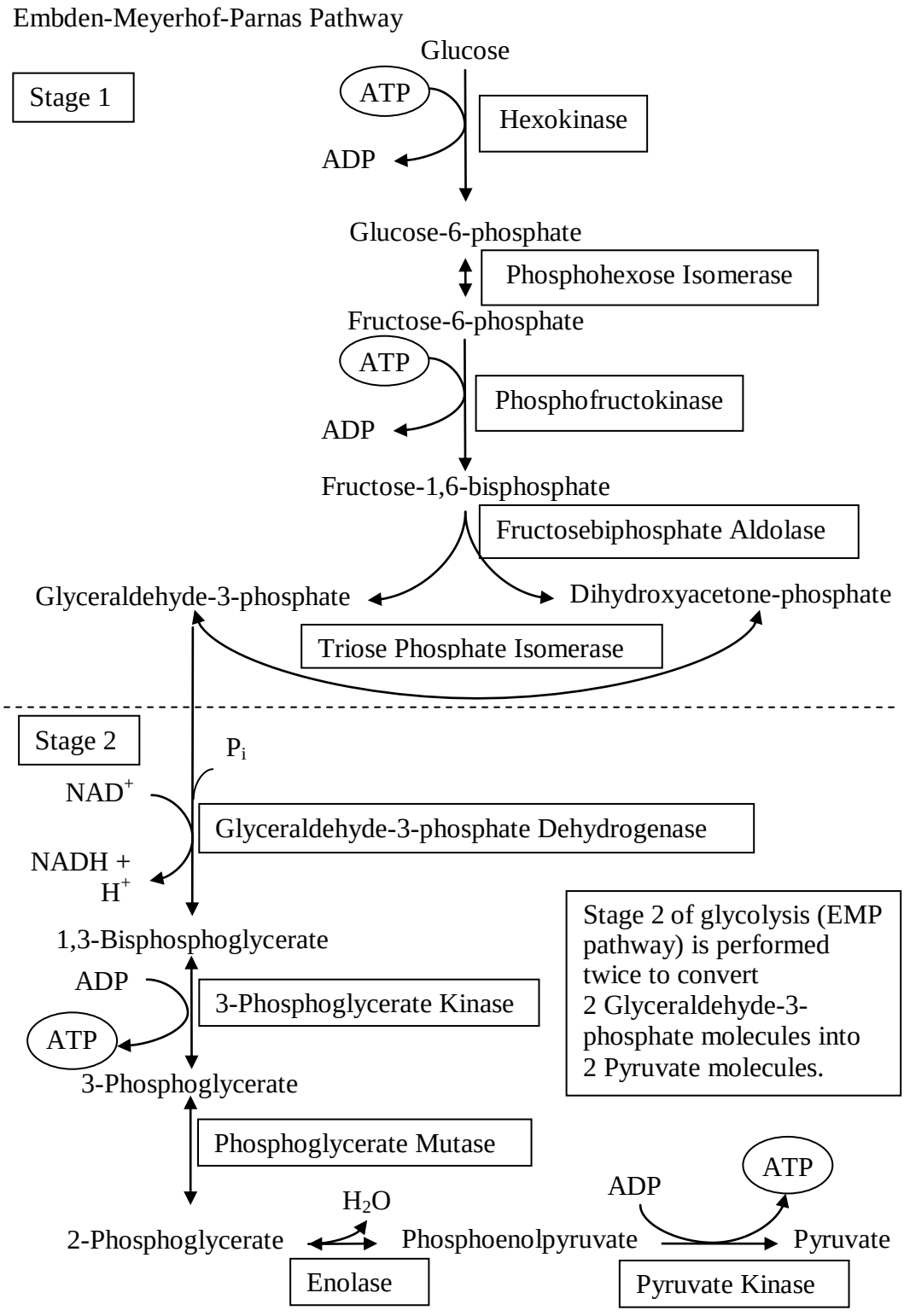


Figure 2.1 Embden-Meyerhof-Parnas Pathway (Glycolysis)

The conventional phosphorylated Entner-Doudoroff (ED) pathway is illustrated in Figure 2.2. The first 4 steps of this pathway produce pyruvate and glyceraldehyde-3-phosphate and require 1 ATP molecule. The glyceraldehyde-3-phosphate is then converted to one pyruvate molecule and 2 ATP molecules via one cycle of the second stage of the EMP pathway, with a net yield of 1 ATP. The non-phosphorylated ED pathway has a net ATP yield of zero and is found in some archaea species (Schönheit and Schäfer 1995; Selig et al. 1997). Selig et al. (1997) used two methods, enzymatic assays for pathway-specific enzymes and ^{13}C-labeling with NMR analysis, to determine which of these three pathways are active in *T. maritima*. Enzymatic assays revealed the presence of all EMP enzymes and the KDPG aldolase present in the phosphorylated ED pathway, but the KDG aldolase from the non-phosphorylated ED pathway was not present. Labeling studies with [1-^{13}C]glucose and [3-^{13}C]glucose showed that ~87% of glycolysis occurred via the EMP pathway and ~13% through the phosphorylated ED pathway (Selig et al. 1997).

T. maritima growth rates are significantly slower when monosaccharides are provided as substrates compared with growth on related polysaccharides (Chhabra et al. 2003). This could possibly be due to the fact that without a PTS glucose requires 1 ATP for phosphorylation prior to glycolysis (Hansen and Schonheit 2003), whereas the cellobiose phosphorylase appears to be a mechanism to conserve ATP, for every 2 molecules of glucose that are available in one cellobiose molecule, one is phosphorylated without the use of ATP, so the ATP requirement for growth on cellobiose-based materials should be half as much as for growth on glucose alone.

It is of interest that although there is activity on mannan, the mechanism by which mannose is incorporated into glycolysis is unclear, because *T. maritima* ATP-dependent glucokinase does not phosphorylate mannose (Hansen and Schonheit 2003; Conners et al. 2006). Mannose has been shown to give an extremely low H_2 yield in

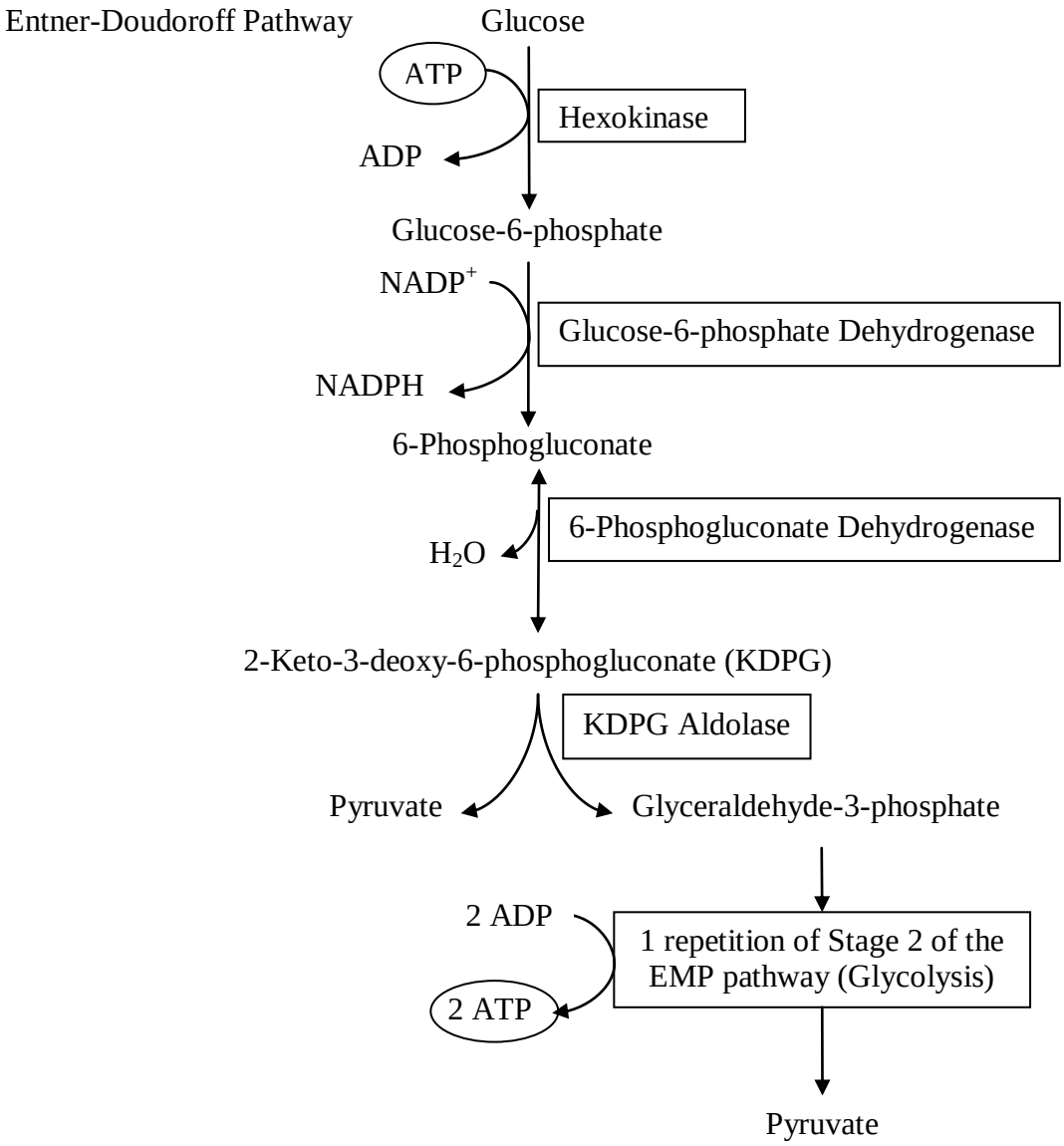


Figure 2.2 Entner-Doudoroff Pathway

T. maritima (Woodward et al. 2002), so it would not be an ideal substrate for biological hydrogen production, although from a physiological standpoint mannose metabolism is clearly of interest.

Connors et al. (2006) speculate that mannose is converted to mannitol or mannitol-6-phosphate and then isomerized into fructose or fructose-6-phosphate, and that NADH demands for these reactions would explain the decrease in H_2 yields. It is also unclear how marine bacteria such as *Thermotoga* species would encounter mannans, which are usually found in plant hemicelluloses (Parker et al. 2001). Parker et al. (2001) speculate that mannan degradation is present in *Thermotoga* species in order to degrade the mannan-based exopolysaccharide material of other hyperthermophiles that may be present in the environment.

2.2.4.2 Pentose Phosphate Pathway

Thermotoga species are capable of metabolizing pentoses, including xylose, arabinose, and ribose (Conners et al. 2005). Following transport of pentoses into the cell, the monosaccharides are phosphorylated. The reactions are shown in Figure 2.3a, both arabinose and xylose are isomerized prior to phosphorylation, but ribose is directly phosphorylated. Xylose isomerase was cloned from *T. neapolitana* (Vielle et al. 1995) and has been the subject of directed evolution; because it has commercial applications for the production of high fructose corn syrup (Sriprapundh et al. 2003).

The phosphorylated pentoses are converted into phosphorylated hexoses by transketolase and transaldolase as illustrated in Figure 2.3b. Two phosphorylated pentoses yield 2 fructose-6-P molecules and 1 glyceraldehyde-3-P molecule. Transketolase and transaldolase are non-oxidative enzymes of the pentose phosphate

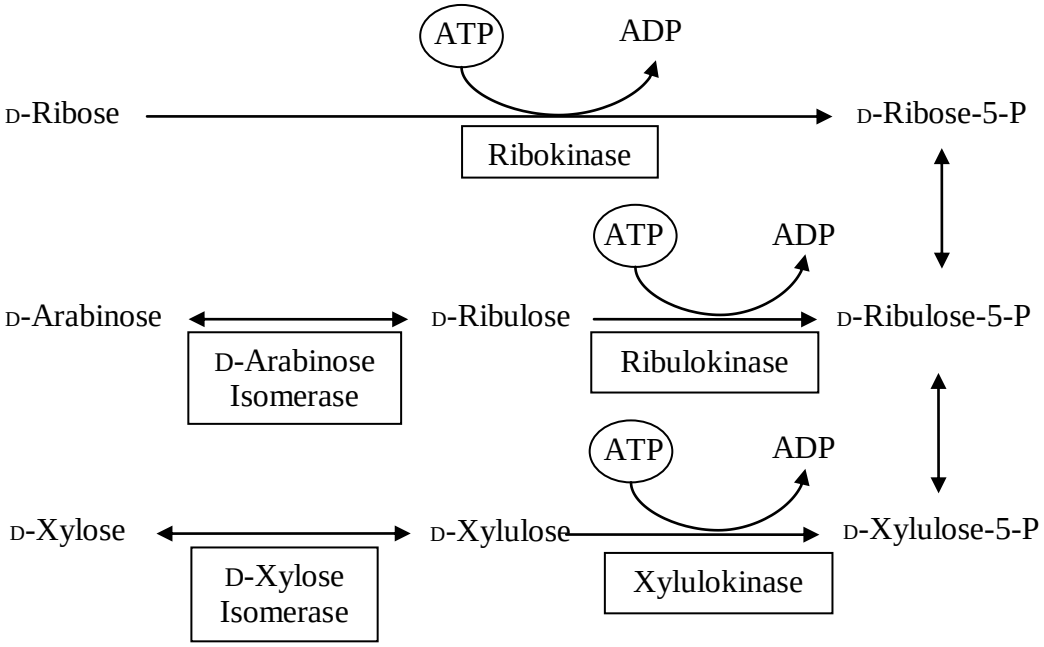
pathway. Genes encoding oxidative pentose phosphate pathway enzymes have been identified in the *T. maritima* genome, but are inactive under experimental conditions used to date (Connors et al. 2006, Nelson et al. 1999).

2.2.5 Hydrogen Formation

Pyruvate is fermented to several products in *T. neapolitana* as indicated in Figure 2.4 (Schröder et al. 1994; Eriksen et al. 2008; Munro et al. 2009). The majority of the pyruvate (>>2 mol) produced from glucose is used for production of approximately 2 mol acetate, 2 mol CO_2, and 2 mol H_2 via the phosphoroclastic reaction. An additional 2 mol of H_2 can be produced by substrate-level phosphorylation during glycolysis; however the actual amount is dependent on biosynthetic requirements for NAD^+ availability and other limitations such as H_2 partial pressure.

It has recently been discovered that both *T. maritima* and *T. neapolitana* have a trimeric bifurcating hydrogenase that can synergistically utilize NADH and Fd to generate H_2 (Schut and Adams 2009). Before identification of this synergistic activity, it was hypothesized that an intermediate NADH:Fd oxidoreductase transferred electrons from NADH to reduce ferredoxin. The reduced ferredoxin (Fd-red) from substrate phosphorylation and the phosphoroclastic reaction could then act with the Fd-only monomeric hydrogenase that is found in both *T. maritima* (TM0201) and *T. neapolitana* (CTN_0485). However, in experimental work *T. maritima* expressed the trimeric hydrogenase and when it was purified it did not interact with Fd or NADH *in vitro* (Verhagen et al. 1999; Jenney and Adams 2008). However when Fd and NADH were tested simultaneously, both were utilized and the question of

A) Pentose phosphorylation



B) Non-oxidative pentose phosphate pathway

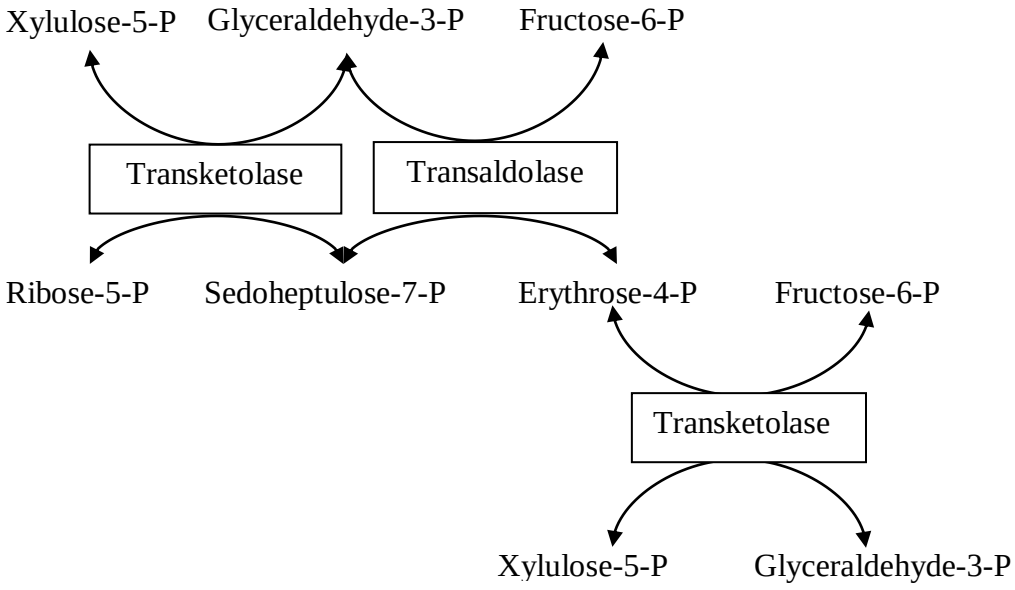


Figure 2.3 Pentose incorporation into glycolysis. A) pentose phosphorylation followed by B) non-oxidative pentose phosphate pathway reactions, conversion of phosphorylated pentoses to fructose-6-P and glyceraldehyde-3-P.

which electron carrier could interact with the trimeric hydrogenase was resolved (Schut and Adams 2009).

Both the trimeric and monomeric hydrogenases in *T. maritima* and *T. neapolitana* are [FeFe]-hydrogenases (also referred to as Fe-only) (Juszczak et al. 1991; Verhagen and Adams 2001; Tosatto et al. 2008). The monomeric hydrogenase is similar to those found in *Clostridia* sp., whereas the trimeric form has only been identified biochemically in *T. maritima* although it is predicted to be in other species based on genome annotations (Schut and Adams 2009). The trimeric structure was determined by Verhagen et al. (1999). The α-subunit has the catalytic activity for conversion of H^+ to H_2, the β-subunit is hypothesized to be the site of NADH oxidation and the γ- subunit is hypothesized to oxidize ferredoxin (Verhagen et al. 1999; Schut and Adams 2009). It was determined by Schut and Adams (2009) that the trimeric hydrogenase utilizes NADH and Fd in 1:1 ratio. Due to biosynthetic requirements for NAD^+ an additional reaction is necessary to regenerate NAD^+ from NADH, Schut and Adams propose that a membrane-bound ion-translocating NADH:Fd oxidoreductase could be responsible for this reaction and are experimentally testing this hypothesis (2009). In Chapter 4 a Na^+-translocating NADH:Fd oxidoreductase is included in an *in silico* model of *T. neapolitana* and model simulations predict cell growth. Experimental evidence for membrane oxidoreductase activity in *T. neapolitana* was found by Kaslin et al. (1998). Both hydrogenase and oxidoreductase activity were found in the cell membrane fraction of *T. neapolitana* following ultracentrifugation. Using membrane vesicles prepared from *T. neapolitana* membranes, it was determined that the oxidoreductase activity was on the cytoplasmic face of the membrane vesicles whereas the hydrogenase activity was on the extracytoplasmic face (Käslin et al. 1998).

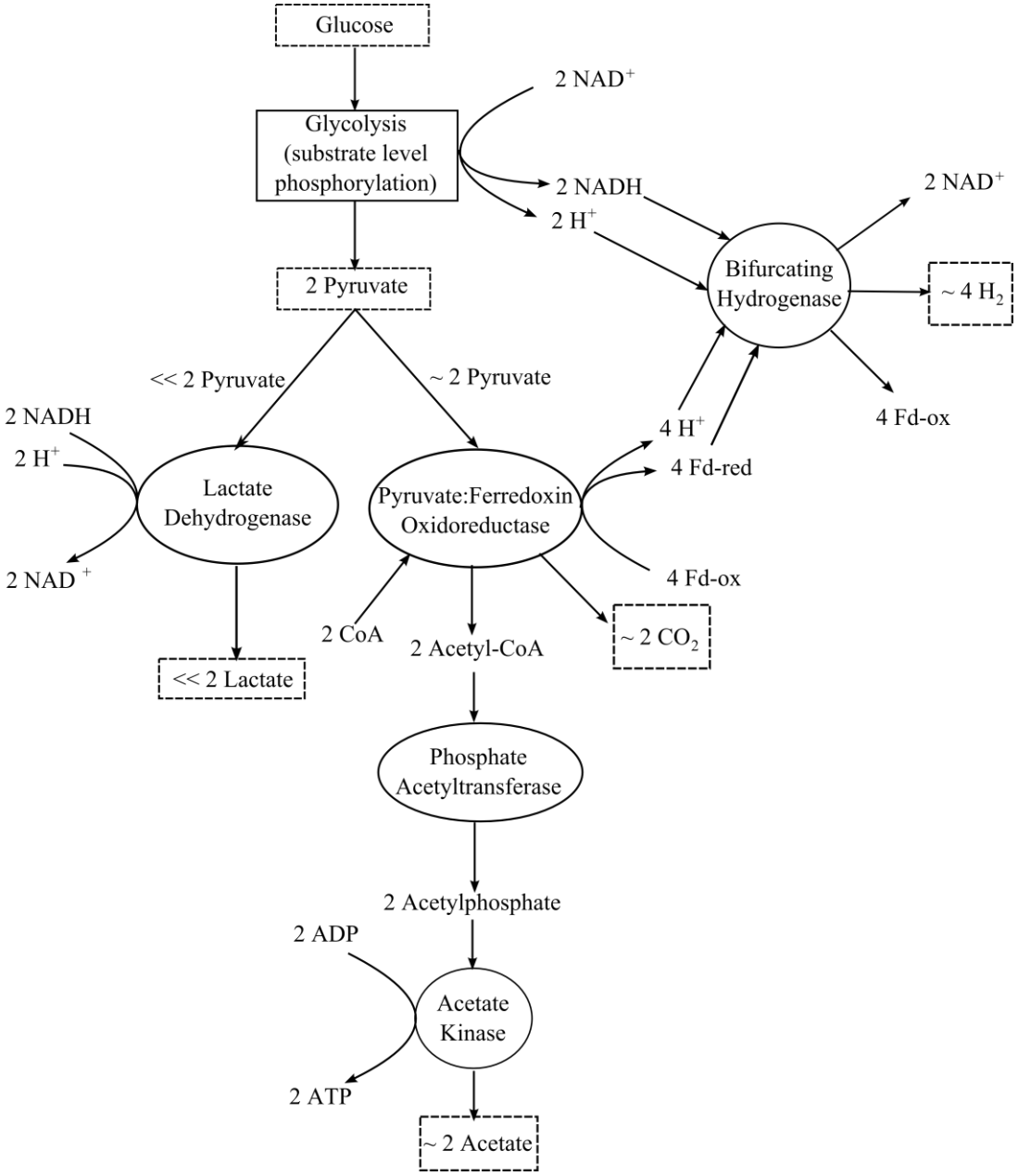


Figure 2.4 Pyruvate metabolism and hydrogen formation in *T. neapolitana*. The majority of the pyruvate is used to produce CO_2, acetate and H_2, a small fraction goes to produce lactate, when the available NAD^+ pool is depleted. H_2 is also generated from substrate-level phosphorylation during glycolysis. The newly identified bifurcating hydrogenase synergistically uses NADH and ferredoxin (Fd) to generate H_2.

2.2.6 Amino Acid and Hydrogen Sulfide (H_2S) Formation

As mentioned previously, L-alanine is also produced by *Thermotoga* species (Ravot et al. 1996), however it was not included in our carbon balance on *T. neapolitana,* because a closed balance was obtained without L-alanine quantification and the expected amount would have been negligible (Munro et al. 2009).

T. neapolitana and *T. maritima* also have sulfur reduction capabilities, but sulfur is not required for growth. When a source of inorganic sulfur is present in the medium, H_2S is evolved. Childers and Noll (1994) determined that a NADH:polysulfide oxidoreductase is the primary sulfur reductase for *T. neapolitana*. Janssen and Morgan (1992) determined that sulfur reduction was not used for growth of *Thermotoga* sp. strain FjSS3.B1, as a detoxification reaction, or energy-generating mechanism, but instead operates as an electron sink that is energetically more favorable than H_2 to replenish intracellular electron carrier pools. For biological hydrogen production purposes, sources of inorganic sulfur should be excluded from the growth medium, to prevent the production of H_2S and subsequent decrease in H_2 production.

2.3 Existing Mathematical Models of H_2 Fermentation

Many experimental studies on biological hydrogen production report important process parameters such as H_2 yield (mol H_2/mol substrate) and volumetric productivity rates. Comparatively fewer researchers have applied modeling techniques to understand the process of fermentative H_2 production. Both unstructured and structured models have been applied to H_2 fermentation processes.

Unstructured models treat cell biomass as a single "black-box" entity and are useful to establish basic process parameters. Structured models divide a cell into distinct components; many structured models include definitions of metabolic pathways (Bailey and Ollis 1986; Nielsen et al. 1991; Shuler and Kargi 2002). If sufficient data is available, structured metabolic models can provide valuable insight for directed genetic manipulation of metabolic pathways.

2.3.1 Unstructured Models

2.3.1.1 Kinetic Models

Several types of unstructured kinetic models have been applied to model biological H_2 production processes. The two most well-known, the Monod and Michaelis-Menten equations (eq 2.18 and 2.19), are used to describe growth and substrate utilization rates:

$$\mu = \frac{\mu_{max} S}{K_S + S} \tag{2.18}$$

$$v = \frac{v_{max} S}{K_S + S} \tag{2.19}$$

where:

K_S = half-saturation substrate concentration ($g \cdot L^{-1}$),

S = substrate concentration ($g \cdot L^{-1}$),

μ = growth rate (h^{-1}),

μ_{max} = maximum growth rate (h^{-1}),

v = substrate utilization rate (h^{-1}), and

v_{max} = maximum rate of substrate utilization (h^{-1}).

The Monod model directly relates substrate availability and growth rate, and the model parameters, μ_{max} and K_S, can be useful for process design. The Monod model requires the assumption that one particular substrate is limiting to growth, once that substrate is depleted growth will cease. The Monod model does not account for endogenous decay or inhibition that could significantly impact growth. Kinetic constants for Monod models of growth in H_2 fermentations are presented in Table 2.3. There was wide variability between studies for estimations of K_S and μ_{max}.

Kinetic constants for the Michaelis-Menten model are presented in Table 2.4. Interpretations of the Michaelis-Menten parameter, v_{max}, varied between studies depending on whether v_{max} was used to describe the maximum rate of H_2 formation (Lo et al. 2008, Chen et al. 2006) or the maximum rate of substrate utilization (Mu et al. 2006). Sufficient data was not presented in these studies for the conversion of the v_{max} values to common units.

Using the Monod and Michaelis-Menten parameters, Lo et al. (2008) attempted to draw comparisons between two different *Clostridia* species grown on two different substrates, but there is no common ground for such a comparison. Therefore, the comparisons Lo et al. (2008) made in their study are not valid. Chen et al. (2006) could compare substrate utilization kinetics of different carbohydrate substrates: sucrose, non-fat dry milk (NFDM), and food waste, because they used the same inoculum source on all the substrates. The K_S values from the Michaelis-Menten equation correlate to substrate affinity; a low K_S value indicates high affinity. Sucrose had the lowest K_S value and the highest carbohydrate content of the substrates, thus Chen et al. (2006) concluded that substrate affinity increases with increased carbohydrate content. However, the maximum H_2 production rate was found on the substrate with the lowest carbohydrate content (Chen et al. 2006), which could indicate a trade-off between carbohydrate content and H_2 production rate or it could be

Table 2.3 Kinetic constants derived from Monod models of growth in H_2 fermentations

Organism(s)	Fermentation Type	Substrate	K_S (g COD·L^{-1})	μ_{max} (h^{-1})	Reference
Clostridium butyricum CGS5	Batch	Xylose	0.67	0.15	Lo et al. (2008)
Clostridium pasteurianum CH4	Batch	Sucrose	4.39	0.31	Lo et al. (2008)
Enterobacter cloacae DM11	Batch	Glucose	5.88	0.40	Nath et al. (2008)
Mixed Culture	Continuous	Sucrose	0.068	0.172	Chen et al.(2001)
Mixed Culture	Continuous	Glucose	0.21	0.75	Whang et al. (2006)
		Glucose with decay	0.19	1.00	
		Peptone	3.07	0.80	
Caldicellulosiruptor saccharolyticus	Batch	Sucrose	0.801	0.130	van Niel et al. (2003)*

* Substrate inhibition was included in the model in this study

Table 2.4 Kinetic constants derived from Michaelis-Menten models for the rate of H_2 formation or substrate utilization.

Organism(s)	Fermentation Type	Substrate	K_s (g COD·L^{-1})	v_{max}	Reference
Clostridium butyricum CGS5	Batch	Xylose	0.5	150 mL H_2·(g vss·h)$^{-1}$	Lo et al. (2008)
Clostridium pasteurianum CH4	Batch	Sucrose	3.7	290 mL H_2·(g vss·h)$^{-1}$	Lo et al. (2008)
Mixed Culture	Batch	Sucrose	1.446	13.9 mL H_2·h^{-1}	Chen et al. (2006)
		NFDM	6.616	25.6 mL H_2·h^{-1}	
		Food Waste	8.692	29.9 mL H_2·h^{-1}	
Mixed Culture	Batch	Sucrose	15.2	0.31 g COD·(g vss·h)$^{-1}$	Mu et al. (2006)

due to the complexity of the NFDM and food waste substrates, which contain other compounds that might enhance H_2 production.

The Monod model is useful for describing the relationship between a limiting substrate and biomass formation, however it does not evaluate the relationship between biomass and product formation. The Luedeking-Piret model (Luedeking and Piret 1959; Wang et al. 1979) has been established to describe the association between product formation and growth:

$$\frac{dP}{dt} = \alpha \frac{dX}{dt} + \beta X \tag{2.20}$$

where:

dP/dt = product formation rate (g product$\cdot L^{-1} \cdot h^{-1}$),

dX/dt = biomass formation rate (g biomass$\cdot L^{-1} \cdot h^{-1}$),

X = biomass concentration (g biomass$\cdot L^{-1}$),

α= growth associated term (g product$\cdot$g biomass^{-1}), and

β = non-growth associated term (g product$\cdot$g biomass$^{-1} \cdot h^{-1}$).

This model can be rewritten to define q_P, the specific productivity rate (g product$\cdot$g biomass$^{-1} \cdot h^{-1}$):

$$\frac{1}{X}\frac{dP}{dt} = q_P = \alpha\mu + \beta \tag{2.21}$$

The Luedeking-Piret model accounts for three possible product formation scenarios; that growing cells produce a product in proportion with growth (the α term), that product formation is proportional to the concentration of cells regardless of growth phase (the β term), or that product formation is mixed, with both growth and non-growth associations (Wang et al. 1979; Bailey and Ollis 1986). The Luedeking-Piret equation is a useful model, because it accounts for the possibility that product formation may have both non-growth and growth-associated components, and allows for estimation of the contribution of both terms to the total product formation. This

mixed-model situation is often the case in energy-yielding metabolism such as anaerobic fermentations, wherein the α term is associated with energy used for growth and the β term is associated with the energy used for maintenance (Bailey and Ollis 1986). Mu et al. (2006) applied the Luedeking-Piret model to experimental data and determined that all three products, acetate, butyrate, and hydrogen were growth-associated; the model results estimated the β terms for all products to be equal to zero. Lo et al. (2008) also used the Luedeking-Piret model to calculate the specific H_2 production rate; however Lo et al. (2008) made the initial assumption that the non-growth term, β, was negligible. Mu et al. (2006) and Lo et al. (2008) estimated the goodness of fit of the Luedeking-Piret model to the experimental data with the coefficient of determination, R^2, defined as the proportion of the total variability in the data that is accounted for by the regression model (Vining 1998).

Kinetic models have been applied to continuous fermentations for H_2 production in two studies (Chen et al. 2001; Whang et al. 2006) and the Monod kinetic parameters from both are in Table 2.3. Chen et al. (2001) derived steady state mass balances (Bailey and Ollis 1986) to describe biomass formation, substrate degradation and product formation:

$$\frac{dX}{dt} = 0 = (\mu - D)X \tag{2.22}$$

$$\frac{dS}{dt} = 0 = D(S_o - S) - \frac{1}{Y_{X/S}}\mu X \tag{2.23}$$

$$\frac{dP_{L,i}}{dt} = 0 = -DP_{L,i} + q_{P_{L,i}} X \tag{2.24}$$

$$\frac{dP_{G,i}}{dt} = 0 = -D'P_{G,i} + q_{P_{G,i}} X \tag{2.25}$$

where:

X = biomass concentration (g biomass·L^{-1}),

S = substrate concentration (g·L^{-1}),

S_o = initial substrate concentration ($g \cdot L^{-1}$),

$P_{L,i}$ = concentration of the i^{th} liquid product ($g \cdot L^{-1}$),

$P_{G,i}$ = concentration of the i^{th} gas product ($g \cdot L^{-1}$),

D = liquid phase dilution rate (h^{-1}),

D' = gas phase dilution rate (h^{-1}),

$Y_{X/S}$ = biomass yield coefficient (g biomass/g substrate)

$q_{P_{L,i}}$ and $q_{P_{G,i}}$ are the specific product formation rate of the i^{th} liquid product and the i^{th} gas product ($g \cdot g\ biomass^{-1} \cdot h^{-1}$).

Chen et al. (2001) assumed that product formation was growth associated, and used a simplified form of the Luedeking-Piret model:

$$q_{P_{L,i}} = \alpha_{P_{L,i}} \mu \tag{2.26}$$

$$q_{P_{G,i}} = \alpha_{P_{G,i}} \mu \tag{2.27}$$

Chen et al. (2001) used eq 2.26 and 2.27 and the assumption that the growth rate could be represented by the Monod equation (eq 2.18) to obtain the following steady state expressions for S, X, $P_{L,i}$, $P_{G,i}$, and r_{H_2} (the production rate of H_2, $g \cdot (L \cdot h)^{-1}$):

$$S = \frac{DK_S}{\mu_{max} - D} \tag{2.28}$$

$$X = Y_{X/S}\left(S_o - \frac{DK_S}{\mu_{max} - D}\right) \tag{2.29}$$

$$P_{L,i} = \alpha_{P_{L,i}} X \tag{2.30}$$

$$P_{G,i} = \alpha_{P_{G,i}}\left(\frac{D}{D'}\right)X \tag{2.31}$$

$$r_{H_2} = D'\alpha_{H_2} X \tag{2.32}$$

based on their model and experimental results, Chen et al. (2001) showed a positive correlation between dilution rate and r_{H_2}. However as dilution rates increase, biomass concentration decreases; since biomass generates H_2, decreased biomass is not desirable. To address this issue Chen et al. (2001) recommended operation at high

dilution rates with the use of either a cell recycle or immobilization system to maintain high cell density in the reactor despite the high dilution rate. However van Niel et al. (2003) found that inhibition by hydrogen was dependent on culture density, so such a system with high dilution rate and high cell density also requires a high H_2 removal rate.

Whang et al. (2006) expanded upon the steady-state kinetic model of Chen et al. (2001), developing three different kinetic models incorporating dual substrates and endogenous metabolism. The first model has identical equations to those used by Chen et al. (2001) (eq 2.28 and 2.29). An endogenous decay term, k_e (h^{-1}) is added to the second model:

$$S_G = \frac{(D + k_e)K_{SG}}{\left(\mu_G^{max} - D - k_e\right)} \tag{2.33}$$

$$X = \frac{DY_{X/G}\left(S_{G0} - S_G\right)}{(D + k_e)} \tag{2.34}$$

In the third model, peptone metabolism is added to the second model for glucose metabolism with endogenous decay. In this model, S_G (glucose) is represented by equation 2.33 and the equations for S_P (peptone) and X are:

$$S_P = \frac{(D + k_e)K_{SP}}{\left(\mu_P^{max} - D - k_e\right)} \tag{2.35}$$

$$X = \frac{DY_{X/G}\left(S_{G0} - S_G\right)}{(D + k_e)} + \frac{DY_{X/P}\left(S_{P0} - S_P\right)}{(D + k_e)} \tag{2.36}$$

The VFA products for all models, were estimated assuming Luedeking-Piret growth associated product kinetics using the same formula as in the Chen et al. (2001) study (see eq 2.33). However Whang et al. (2006) derived an equation for H_2 production rate based on the stoichiometric relationships between H_2 and the VFA products :

$$P_{H_2} = \left(2 \times \frac{P_{HAc}}{60} + 2 \times \frac{P_{HBu}}{88} - \frac{P_{HFo}}{46} - \beta_{NH_4^+ - N} P_{NH_4^+ - N}\right) \times D \tag{2.37}$$

where:

P_{H_2} = H_2 production rate (mmol $H_2 \cdot L^{-1} \cdot h^{-1}$),

P_{HAc} , P_{HBu} , and P_{HFo} are acetic, butyric, and formic acid ($mg \cdot L^{-1}$),

D = dilution rate (h^{-1}), and

$\beta_{NH_4^+-N} P_{NH_4^+-N}$ (mmol $H_2 \cdot L^{-1}$) represents the effect of consumption of ammonium nitrogen on H_2 production.

it should be noted that the ammonium nitrogen term is not included when the single substrate models are used and the term in parentheses in eq 2.37 is the equivalent of the Luedeking-Piret α term for H_2.

The three models were compared using posterior probability calculations (Box and Hill 1967). The posterior probability of model i being correct is (Whang et al. 2006):

$$Pr_i = \frac{R_i}{\sum R_i} \tag{2.38}$$

where the summation includes a total of i models

$$R_i = (\text{RSS Ratio})^{0.5(n-p)} \tag{2.39}$$

and

$$\text{RSS Ratio} = \frac{RSS_{min}}{RSS_i} \tag{2.40}$$

where:

RSS_i = residual sum of squares of model i,

RSS_{min} = smallest RSS among the models,

n = number of observations, and

p = number of estimated parameters.

This method allows for comparison between distinct models, because it accounts for the degrees of freedom in each model. For predictions of biomass, acetate, butyrate and H_2 production, the third model, with dual substrates and endogenous metabolism, had a 100 % posterior probability of being correct while Models 1 and 2 had 0%

posterior probability of being correct (Whang et al. 2006). For glucose consumption it was calculated that there was a 55% probability for Model 3 being correct and a 45% probability for Model 2. The product concentration for all three models was dependent on the biomass model; therefore it makes sense that the product predictions had similar posterior probabilities to biomass for all the models. The high posterior probability of the third model for biomass can be attributed to the fit of the model to the data at low dilution rates. Model 1 and Model 2 had poor fits to the data at low dilution rates. For the case of Model 1 growth was overestimated at low dilution rates, due to the lack of an endogenous metabolism term. For Model 2 growth was underestimated at lower dilution rates, because peptone was present and consumed in the reactor at these dilution rates, but not accounted for in the model. The dual substrate with endogenous decay model appears to be promising for investigation of the effect of organic nitrogen metabolism on H_2 production in continuous culture, and could also be applied for multiple sugar substrates, i.e. a combination of hexoses and pentoses.

Aceves-Lara et al. (2008) also assumed that growth followed the Monod model in continuous fermentation, but applied a dynamic modeling approach, wherein the stoichiometric reactions of the system are estimated using data obtained from perturbation of a continuous culture system. In this study, the varied parameters were influent glucose concentration and influent flow rate. A mass balance on the reactor system is represented by the following equation where the change in time of the state variable concentrations of glucose, acetate, propionate, butyrate, X (biomass), CO_2, and H_2 are equal to the rate of product formation minus the rates of removal from the system of the liquid and gas products:

$$\frac{d}{dt}\begin{bmatrix} Glucose \\ Acetate \\ Propionate \\ Butyrate \\ X \\ CO_2 \\ H_2 \end{bmatrix} = K \cdot r - D\begin{bmatrix} Glucose - Glucose_{in} \\ Acetate \\ Propionate \\ Butyrate \\ X \\ CO_2 \\ H_2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ q_{CO_2,gas} \\ q_{H_2,gas} \end{bmatrix} \tag{2.41}$$

where:

K = matrix of pseudo-stoichiometric coefficients,

r = vector representing the growth rate equations based on the Monod model, with adjusted units $(g\cdot(L\cdot h)^{-1})$,

D = dilution rate (h^{-1}), and

$q_{CO_2,gas}$ and $q_{H_2,gas}$ are the effluent gas flow rates of CO_2 and H_2 $(g\cdot(L\cdot h)^{-1})$.

At steady state equation 2.41 reduces to:

$$K \cdot r = U \tag{2.42}$$

where U is defined as

$$U = D\begin{bmatrix} Glucose - Glucose_{in} \\ Acetate \\ Propionate \\ Butyrate \\ X \\ CO_2 \\ H_2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ q_{CO_2,gas} \\ q_{H_2,gas} \end{bmatrix} \tag{2.43}$$

and K and r are defined as:

$$K = \begin{bmatrix} -K_{11} & -K_{12} \\ 0 & K_{22} \\ K_{31} & K_{32} \\ K_{41} & K_{42} \\ K_{51} & K_{52} \\ K_{61} & K_{62} \\ K_{71} & 0 \end{bmatrix} \tag{2.44}$$

$$r = \begin{bmatrix} \dfrac{\mu_{max,1}\,Glucose}{K_{Glu1} + Glucose} \\ \dfrac{\mu_{max,2}\,Glucose}{K_{Glu2} + Glucose} \end{bmatrix} X \tag{2.45}$$

the size of K is determined by principle component analysis of U (Bernard and Bastin 2005; 2005). Concentration data is obtained from experimental perturbations of the continuous fermentation system, this data is then used to generate several vectors, u(t) that can be combined to form U. The singular values of U are computed to determine the number of reactions that should be represented by the columns of K. All singular values of U will be non-zero, so there will be a large number of possible reactions in K. To practically determine the minimum number of reactions in K, it is recognized that each singular value represents the weight that the associated eigenvector (reaction) would have if included in the K matrix. Aceves-Lara et al. (2008) chose a 95% confidence interval and found that a minimum of 2 reactions in K were enough to satisfy this degree of variance.

Once the number of reactions in the network has been defined the values of the K_{ij} reaction components can be determined using optimization techniques with the U and K matrices. The kernel (null space) of K^T is obtained with the constraint that $0 \leq K_{ij} \leq 1$ and the minimum numerical values of K are obtained by the following formula:

$$\min(Ker(K^T)^T U) \tag{2.46}$$

The resulting matrix K showed that reaction 1 was associated with hydrogen production and reaction 2 was associated with acetate production. However, this result requires more investigation and these reactions could not be associated with specific members of the bacterial community, because the pseudo-stoichiometric coefficients for acetate and H_2 were forced to be zero in reaction 1 and 2, respectively, to meet the mathematical requirements of the model. However, Aceves-Lara et al.

(2008) did attempt to show robustness of the model by running a second set of perturbation experiments and fitting the data with the model from the first experiment, and found that there was good agreement between the model and the second data set.

Models with inhibition terms have also been applied to H_2 production studies. It is important to include inhibition, because it allows for more accurate parameter estimation. For example, μ_{max} can be underestimated in a system if inhibiting compounds are present, but not accounted for in the model. Andrews (1968) adapted the Haldane model for enzyme kinetics with inhibition (Haldane 1930) to describe microbial growth kinetics with inhibition:

$$\mu = \frac{\mu_{max} S}{K_S + S + \frac{S^2}{K_I}} \tag{2.47}$$

where

μ = specific growth rate (h^{-1}),

μ_{max} = maximum specific growth rate (h^{-1}),

K_s = half-saturation substrate concentration ($g \cdot L^{-1}$), the lowest substrate concentration at which the specific growth rate is equal to one-half the maximum specific growth rate in the absence of inhibition,

K_I = substrate inhibition constant ($g \cdot L^{-1}$), the highest substrate concentration at which the specific growth rate is equal to one-half the maximum specific growth rate in the absence of inhibition, and

S = substrate concentration ($g \cdot L^{-1}$).

Nath et al. (2008) compared the Andrews model with the Monod equation to model growth in the initial dark fermentation stage of a two-stage fermentation process. This was done to determine if substrate inhibition could account for a small difference between the experimental data and the Monod model prediction. Nath et al. (2008) did not publish the kinetic parameters for the Andrews model, so the results of

the Andrews and Monod models cannot be compared. Nath et al. (2008) did however make the claim that based on statistical analysis of the data and the Andrews model predictions for both biomass formation and substrate consumption, the null hypothesis could not be rejected. This indicates that there was no significant difference between the Andrews model and experimental results, thus Nath et al. (2008) claimed that the inhibition model accurately described the data.

Another inhibition model based on the Monod equation developed by Han and Levenspiel (1988) was used by van Niel et al. (2003) and Zheng and Yu (2005):

$$r = r_{max}\left(1-\frac{C}{C_{crit}}\right)^{n}\left(\frac{S}{S+K_s}\right) \tag{2.48}$$

where:

r = H_2 production rate (mL·g biomass^{-1}·L^{-1}),

r_{max} = maximum H_2 production rate (mL·g biomass^{-1}·L^{-1}),

C = inhibitor concentration (g·L^{-1}),

C_{crit} = critical inhibitor concentration at which H_2 production ceases (g·L^{-1}),

n = degree of noncompetitive inhibition,

K_s = apparent half velocity constant for the substrate (g·L^{-1}), and

S = substrate concentration (g·L^{-1}).

When substrate concentration was not limiting ($S >> K_s$) equation 48 reduces to:

$$r = r_{max}\left(1-\frac{C}{C_{crit}}\right)^{n} \tag{2.49}$$

One advantage that the Han-Levenspiel model offers over other inhibition models is that the critical inhibitor concentration, C_{crit}, defined as the concentration at which the production rate ceases, can be directly estimated. Other inhibition models do not predict this value (Han and Levenspiel 1988). Another advantage of this model over substrate inhibition models, such as the Andrews model, is that both substrate and

product inhibition can be represented; the model is not limited to substrate inhibition. Using this model, van Niel et al. (2003) examined inhibition effects of H_2, sucrose, sodium acetate, and sodium chloride in cultures of *Caldicellulosiruptor saccharolyticus*. Sucrose had K_S = 2.09 mM, n = 1.39, and a critical sucrose concentration = 292 mM. Inhibition by salts such as sodium acetate and sodium chloride were determined to be related to ionic strength for two reasons. First the inhibition kinetics of sodium acetate and sodium chloride were determined to be identical and secondly, when tested at 2 different pH values there was no significant difference in critical acetate concentration. It was also found that accumulation of H_2 was the greatest inhibitor of H_2 production. H_2 inhibition increased with increased cell density, the critical concentration of H_2 in the headspace was 27.74 mM during the lag phase, 25.12 mM during the early exponential phase, and 17.28 mM during the mid-exponential phase (van Niel et al. 2003). Zheng and Yu (2005) also used the modified Han-Levenspiel equation (eq 49) to evaluate the inhibition effects of butyrate addition for mixed culture H_2 production. The maximum hydrogen production rate was determined to be 59.3 mL $H_2 \cdot (\text{g VSS} \cdot \text{h})^{-1}$, the critical added butyrate concentration was 25.08 g/L, and the degree of inhibition (n) was 0.323. The parameters in both the van Niel et al. (2003) and Zheng and Yu (2005) papers were estimated using non-linear least squares regression, which introduces less error than a linear least squares regression with a linearized form of eq 2.49.

Inhibition was also incorporated into the exponential growth equation by van Niel et al. (2003):

$$X = X_o \exp\left(\mu\left(t - t_o\right)\right) \tag{2.50}$$

where:

X = biomass concentration (g biomass$\cdot L^{-1}$),

t = time (h),

X_o = biomass concentration at time t_o (g biomass·L^{-1}), and

μ (h^{-1}) is defined by the equation:

$$\mu = \mu_{max}\left(1-\frac{C_{acetate}}{C_{acetate,crit}}\right)^{n}\left(1-\frac{C_{sucrose}}{C_{sucrose,crit}}\right)^{m}\left(\frac{S}{S+K_s}\right) \quad (2.51)$$

the μ_{max} obtained using eq 2.51, 0.13 h^{-1}, was double the μ_{max} predicted without accounting for acetate and sucrose inhibition, 0.073 h^{-1}. This result reinforces the idea that inclusion of inhibition terms in models allows for more accurate parameter estimations.

2.3.1.2 Statistical Models

Statistical models can be derived from factorial design and response surface methodology studies wherein multiple test variables are changed simultaneously to optimize a selected response variable and multivariate linear regression is employed to parameterize the system. This strategy differs from the traditional "one-at-a-time" approach, because it allows for not only observing the effect of a specific test variable on the response variable, but also to determine the interaction effects of test variables on the response variable (Montgomery 2005). These studies primarily involve mixed culture systems, wherein the bacterial community is treated as a black box, although it is often noted in these studies that *Clostridia* species dominate (Lay 2000; Lay et al. 2005). Statistical models are particularly useful for H_2 producing mixed communities, because tight control of process parameters is crucial for maintaining the activity of H_2 producing bacteria. Pure culture response surface modeling studies for H_2 formation are not as common as mixed culture studies, although one has been conducted with *Enterobacter cloacae* DM11, a double mutant lacking both alcohol and organic acid formation pathways (Sen and Das 2005). The inoculum sources for mixed culture

studies include sludge (Lay 2000; Lay 2001; Lin and Lay 2004; Mu et al. 2007), compost (Van Ginkel et al. 2001; Lay et al. 2005), and soil (Van Ginkel et al. 2001; Van Ginkel and Logan 2005). A variety of substrates have been employed in these studies including starch (Lay 2000), microcrystalline cellulose (Lay 2001), sucrose (Van Ginkel et al. 2001; Lin and Lay 2004; Lay et al. 2005; Mu et al. 2007), glucose (Sen and Das 2005; Van Ginkel and Logan 2005), and model food waste (dog food) (Lay et al. 2005). A survey among these studies reveals similar test variables that are of interest: initial pH (Lay 2000; Van Ginkel et al. 2001; Sen and Das 2005; Mu et al. 2007), hydraulic residence time (HRT) (Lay 2000; Van Ginkel and Logan 2005), initial substrate concentration (Lay 2001; Van Ginkel et al. 2001; Sen and Das 2005; Van Ginkel and Logan 2005), and the initial substrate to biomass ratio (S_o/X_o) (Lay 2001; Mu et al. 2007). In general these studies focus on using both the H_2 yield and the production rate as the response variables for optimization, with the exception of the Mu et al. (2007) study where maximum growth rate was the response variable. Response surface modeling would be useful for optimizing environmental parameters for H_2 yields and production rate, but reaching and exceeding the theoretical yield of 4 mol H_2 /mol glucose will most likely require metabolic engineering using structured cell models.

2.3.2 Structured Models

2.3.2.1 Metabolic Models

In studies with defined cultures, the creation of structured, metabolic models becomes a possibility. More detail on construction of metabolic models will be provided in section 2.4. Here the existing metabolic flux analysis studies pertaining to

H_2 production will be discussed briefly. Metabolic flux analysis determines metabolic pathway fluxes (also called reaction rates), by construction of a dynamic mass balance equation for the metabolic pathways of interest:

$$\frac{dx}{dt} = S \cdot v \tag{2.52}$$

where:

x = vector of metabolite concentrations (m x 1) (mass·volume^{-1}),

dx/dt = time derivative of the concentration vector (mass·volume^{-1}·time^{-1}),

S = stoichiometric matrix of m metabolites and n reactions (m x n), and

v = flux vector (n x 1) (mass·volume^{-1}·time^{-1}).

Pseudo-steady state is assumed, this means that there is no accumulation of intracellular metabolites (Stephanopoulos et al. 1998), therefore dx/dt is equal to zero and eq 2.52 reduces to:

$$S \cdot v = 0 \tag{2.53}$$

To determine the flux distribution, v, first the degrees of freedom (F) in the system of linear equations is calculated from the equation:

$$F = n - m \tag{2.54}$$

where:

n = number of reactions and

m = number of metabolites in the system.

For biological systems n is always greater than m, and the degrees of freedom, F, represents the number of fluxes that need to be measured to find an exact solution to the system of equations. If exactly F fluxes can be measured, then the system is determined, if less than F fluxes can be measured the system is underdetermined and linear optimization of an objective function is employed to find a space of possible solutions (Stephanopoulos et al. 1998).

Two studies with specific intention to optimize H_2 yields using metabolic flux analysis have been published (Manish et al. 2007; Oh et al. 2008). A third study by Zhang and colleagues (2009) was conducted to develop a model of *T. maritima* central carbon metabolism with integration of *T. maritima* protein crystal structures. This model was used to examine questions related to the evolution of protein structure and function, but did not emphasize the H_2 production potential of *T. maritima*. Tests of the model's capacity to produce H_2 from a variety of carbon sources were included in the supplementary information, but no genetic manipulation of the model or the organism was conducted to either add or delete pathways to improve H_2 production. Likely this is due to the lack of a robust genetic transfer system for the *Thermotogales*.

Manish et al. (2007) created a metabolic model of *E. coli* and an *ldhA* (lactate dehydrogenase) deletion mutant, with the purpose of optimizing H_2 production. In this study metabolic flux analysis was used to solve for intracellular fluxes using experimental data (extracellular fluxes) from batch experiments conducted by Kabir et al. (2005). Enough measured fluxes were available to determine exact solutions for the intracellular flux distributions. The hydrogen production rate was predicted from the model, because it was not experimentally determined in the original study (Kabir et al. 2005). Manish et al. (2007) determined that the wild type *E.coli* strain had a yield of 0.17 mol H_2/mol glucose and the *ldhA* knockout mutant had a yield of 0.23 mol H_2/mol glucose. Since no experimental measurements were made for these predicted values cannot be compared to experimental results. Manish et al. (2007) probably selected this system because the knockout mutant had been prepared, the experimental data was available, and *E. coli* is a widely used model organism with well-defined metabolic pathways.

Oh et al. (2008) conducted batch experiments and created a metabolic model of *Citrobacter amalonaticus* Y19 for optimization of H_2 production. In this study the

system was underdetermined; therefore Oh et al. (2008) employed the linear optimization program MetaFluxNet to solve an objective function for either maximum specific growth rate or maximum specific H_2 production rate. When specific growth rate was maximized, the predicted H_2 yield closely matched the experimentally determined H_2 yield. However, the experimental specific growth rate was actually three times higher than the model value. It is likely that small measurement errors and oversimplified model assumptions are responsible for this discrepancy. Oh et al. (2008) also modified the model network by addition of a non-native hydrogenase to attempt to surpass the maximum yield of 2 mol H_2/mol glucose in facultative anaerobes and found a predicted maximum H_2 yield of 8.47 mol H_2/mol glucose.

2.4 *In Silico* Model Development for *T. neapolitana*

2.4.1 Model Reconstruction and Simulation

Traditionally metabolic engineering had focused on specific metabolic pathways; there has been a paradigm shift from this reductionist approach to the development of holistic modeling approaches to create genome-scale *in silico* models for organisms, integrating the wealth of data available from high-throughput methods such as genomics, proteomics, transcriptomics, metabolomics, and fluxomics. Models have been developed for *Escherichia coli* (Edwards and Palsson 2000), *Haemophilus influenza* (Edwards and Palsson 1999), *Helicobacter pylori* (Schilling et al. 2002), *Saccharomyces cerevisiae* (Forster et al. 2003), and more recently *Geobacter sulfurreducens* (Mahadevan et al. 2006), *Methanosarcina barkeri* (Feist et al. 2006), *Chlamydomonas reinhardtii* (Boyle and Morgan 2009), *Clostridium acetobutylicum* (Senger and Papoutsakis 2008), and *T. maritima* (Zhang et al. 2009). The application

of the systems biology modeling approach to H_2-producing organisms is an area of national interest (Energy 2008).

2.4.1.1 Reconstruction: Building the Stoichiometric (S) Matrix

A biological network can be represented as a system of differential equations as discussed in section 2.3.2.1 (see eq 2.52). The stoichiometric (S) matrix contains the stoichiometry of the biochemical reactions in the network. The columns of S represent the reactions and the rows represent the metabolites in the network. The S matrix is a linear transformation, mapping the vector of fluxes through the network reactions, v, to the time derivatives of metabolite concentrations, dx/dt. The S matrix is created using genomic, biochemical, physiological and literature-based data (Patil et al. 2004; Palsson 2006). S can be genome-scale or smaller, i.e. central carbon metabolism, and can be developed to describe not only metabolism, but also regulatory and signaling networks. The ultimate goal is to integrate all three network types, however only microbial metabolic and transcriptional regulatory networks have been integrated successfully thus far (Herrgård et al. 2004; Palsson 2006).

The process of creating a high-quality metabolic reconstruction is non-trivial and requires extensive time and effort (Thiele and Palsson 2010). In some of our preliminary work constructing a *T. neapolitana* model using a bottom-up approach based on information in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, many inconsistencies and gaps in knowledge became apparent. There is a significant need for both accurate database management and biochemical information related to enzymes and reactions in organisms – the value of this type of data should not be underestimated. Currently ~30 organism models exist and this number should increase as more annotated genomes become available (Thiele and Palsson 2010).

Rather than develop new model reconstructions solely from the primary data sources, a method was developed (see Chapter 4) to create a model for one species using the model of a closely related species. The Mauve genome alignment program was integral to this strategy (Darling et al. 2004), and this method could be applied to more species than the *Thermotogales* species dealt with here.

2.4.1.2 Simulation: Flux Balance Analysis of Constraint-based Models

Once the stoichiometric matrix has been reconstructed, the system of differential equations can be used as a predictive model for phenotypic analysis. In order to produce an exact analytical solution to this system of equations we would need to specify the kinetics of each reaction in the network. Unfortunately most of the available kinetic information in the literature is from *in vitro* experiments conducted under varied conditions, which does not closely approximate *in vivo* enzyme activity (Teusink et al. 2000; Edwards et al. 2002; Patil et al. 2004; Heijnen 2005). Since the evaluation of kinetic parameters in a genome-scale network is presently unrealistic, approximative kinetic modeling formats have been proposed that can be used to obtain an analytical solution to a network (Heijnen 2005). However, approximative kinetic models still have the limitations of requiring the definition of a reference state for the cellular system and the experimental measurement of elasticity coefficients based on perturbations relative to that reference state. Constraint-based modeling is an alternative to these data-driven kinetic modeling approaches. Rather than exactly calculating what a network does, we can narrow the range of possible phenotypes the system can achieve using the application of constraints (Edwards et al. 2002). This modeling approach is referred to as Flux Balance Analysis (FBA), and the major advantage is that it does not require kinetic information. In FBA, linear optimization

is used to find the maximum (or minimum) of a particular objective function (e.g. maximum growth, maximum H_2 production) in the convex solution space defined by the constraints of the model. One disadvantage of FBA is that, because a steady state is assumed, the dx/dt values cannot be calculated; therefore the build-up of toxic intermediates cannot be accounted for in the model (Edwards et al. 2002).

There are four types of constraints that can be applied to the model (Price et al. 2004; Palsson 2006). Physico-chemical constraints include conservation of mass and energy as well as mass-transport and thermodynamic constraints. Spatial/topological constraints include considerations of how the cell is compartmentalized and the high packing density of the intracellular environment. A third category of constraints are environmental constraints, which are time and condition-dependent. These can include availability of electron donors and acceptors, pH, temperature, and osmolarity. For the integration of results from multiple labs it is important that growth conditions and substrates are well-defined. The fourth category, regulatory constraints, is self-imposed and includes transcriptional, translational, and enzymatic regulation mechanisms. Constraints can be defined as balances (equality statements), such as conservation of mass and energy, or as bounds (inequality statements), such as setting minimum and maximum values for each flux in the system.

Becker et al. (2007) created a free MATLAB toolbox that can be used to accomplish constraint-based analysis, referred to as the COBRA (constraint-based reconstruction and analysis) toolbox. Proprietary platforms are also available (e.g. Sympheny^{TM} from Genomatica). Constraint-based models have several applications in addition to predicting the optimum flux distribution for a particular objective function (steady state growth behavior) (Price et al. 2004; Becker et al. 2007). The COBRA toolbox is capable of predicting dynamic growth behavior (batch culture growth), the effects of gene deletions, and can be used for robustness analysis, i.e. the

effect of changing a single flux on the predicted growth rate. Flux variability analysis is another tool in this package, after the optimal solution is found, the range of possible values for each flux to achieve this solution are found (Becker et al. 2007).

Application of additional constraints to the flux balance analysis model can improve the predictive power of the model (Patil et al. 2004; Price et al. 2004). Beard et al. (2002) and Henry et al. (2007) developed methods to apply thermodynamic constraints to flux balance analysis of genome-scale models. The Henry et al. (2007) approach was significantly more comprehensive than the Beard et al. (2002) approach. Beard et al. (2002) eliminated futile cycles from the model by applying a constraint on the entropy of reactions, but this only utilizes the existing values in the S matrix, no additional information was applied to system, whereas Henry et al. (2007) evaluated the thermodynamic feasibility of individual reactions, by calculating the Gibbs free energy of each reaction in the network. The Henry et al. (2007) approach requires the application of group contribution theory, which states that a compound consists of smaller structural subgroups and the Gibbs free energy of formation of these subgroups can be used to calculate the Gibbs free energy of formation of that particular compound. Feist et al. (2007) used the method outlined by Henry et al. (2007) to add thermodynamic constraints to the most recent genome-scale model of *E. coli*.

Improving the integration of regulatory information into these models is an area of active research. The low level of evolutionary conservation of regulatory networks between organisms means that it is difficult to use the comparative genomics approach that works well for development of metabolic networks (Herrgård et al. 2004). However one advantage that regulatory network reconstruction has over metabolic network reconstruction is that genome-wide mRNA expression analysis is an automated and widely-used technique, whereas genome-wide measurements of

metabolite levels and fluxes are not routine or simple procedures (Herrgård et al. 2004). Regulatory information is time-dependent and can be applied to FBA to improve the validity of batch growth predictions. The application of boolean logic operators is described by Covert et al. (2001). For a defined time period of growth, temporary constraints can be applied on fluxes in the network to reflect the presence or absence of the related transcriptional activity, thus qualitative gene expression information can be applied to quantitative flux balance analysis (Covert et al. 2001).

2.4.2 Model Validation

2.4.2.1 Experimental Perturbations

To validate the predictive power of a metabolic model, experimental perturbations must be conducted. There are two possible approaches. Environmental parameters can be manipulated or the activities of key genes or proteins can be specifically modified (Palsson 2006). Environmental parameters might include the carbon and nitrogen substrates, temperature, osmolarity, pH, agitation rate, and gas-phase and liquid-phase dilution rates. Modification of gene and enzyme activity is a more targeted perturbation of the model; however the lack of a robust genetic system for *T. neapolitana* (Noll and Vargas 1997; Yu et al. 2001; Conners et al. 2006) makes this a difficult approach. Modification of enzyme activity could be achieved using specific enzyme inhibitors, small inhibitory RNAs, or the addition of small peptides such as the one found to increase EPS concentration (Johnson et al. 2005).

2.4.2.2 Measurement techniques

In addition to the measurements typically conducted to evaluate fermentations, high-throughput measurement techniques, such as transcriptomics, proteomics, metabolomics, and fluxomics, can be used to iteratively improve and validate a model. Transcriptomics is the analysis of the mRNA present under certain well-defined physiological conditions using DNA microarrays. Many functional genomics studies relying on this method have been conducted in *T. maritima* (section 2.2.6). Transcriptomic analysis is the easiest of the -omic methods to use, however a major limitation of DNA microarray analysis is that it only gives information on transcriptional regulation, not post-transcriptional events such as regulatory activity of proteins. Therefore it is recommended that transcriptome analysis is combined with other methods to provide more detailed results (Dharmadi and Gonzalez 2004). More specifically, Hatzimanikatis et al. (1999) recommend the integration of data from transcriptome and proteome analysis, allowing the complete characterization of regulatory processes. Proteomic analysis is commonly conducted using 2D gel electrophoresis to resolve proteins and the resulting spots are identified as specific proteins using mass spectrometry (MS). Newer methods use LC-MS/MS, wherein the whole proteome is digested with specific proteases and the peptide cocktail is separated with LC methods and run through MS analysis twice (Blomberg 2004).

Just as transcriptomics and proteomics are related, metabolomics and fluxomics are complementary methods. Metabolomics is the quantification of all metabolites inside and outside a cell under defined growth conditions (Mashego et al. 2007). Extracellular metabolites are commonly measured in fermentations; however intracellular metabolites are more difficult to quantify. Fluxomics is the measurement of the intracellular and extracellular fluxes of metabolites, and is accomplished using

^{13}C-labeling of carbon sources. Fluxome analysis is perhaps the most attractive of all the -omic data collection methods, because it is quantitative, directly related to the flux vector of the model, and is considered to be representative of the information that can be obtained from the other -omic analysis methods (Sauer 2004). One major issue with both metabolomic and fluxomic analysis is the need for improvement of rapid sampling and quenching techniques, to ensure that intracellular metabolite concentrations do not change during the sampling phase.

The experimental methods for metabolomic and fluxomic analysis have been extensively reviewed (Mashego et al. 2007; Garcia et al. 2008) and include gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS), capillary electrophoresis-mass spectrometry (CE-MS) and nuclear magnetic resonance (NMR) analysis. According the Sauer (2004) mass spectrometry methods are more useful than the NMR approach, because of sensitivity, decreased cost, and speed, it is estimated that ~300 samples can be processed per day using GC-MS. One drawback of GC-MS analysis is that analytes need to be volatile, and derivatization of analytes is required (Garcia et al. 2008). LC-MS does not require derivatization and has the additional advantage of versatile separation strategies through variation of the mobile and stationary phases. A drawback of LC-MS includes the inability to analyze gas phase metabolites and decreased availability of comprehensive metabolite databases compared with the GC-MS and NMR techniques (Garcia et al. 2008). Fluxomics is the most immediately practical of the 4 -omic methods described here, based on its ease of integration into the FBA model. Regulatory information provided by transcriptomics and proteomics would also be valuable data, considering that regulatory mechanisms in *Thermotoga* species are under active study, although the incorporation of this data into the model is more difficult than integration of flux data.

2.5 Summary

Fermentative thermophilic organisms offer distinct advantages over other production organisms for biological H_2 production (Claassen et al. 1999) and the primary limiting factor for H_2 production is the yield of H_2 from carbohydrate substrates (Nath and Das 2004). The thermophilic bacterium *Thermotoga neapolitana* produces H_2 and although some basic studies on its H_2 production capabilities have been conducted (Van Ooteghem et al. 2002; Eriksen et al. 2008; Nguyen et al. 2008; Munro et al. 2009), a directed approach to improve the H_2 yield in this organism has not been attempted.

Based on biochemical and molecular biology studies in *T. neapolitana* and the related organism *T. maritima* the central carbon metabolic pathways in *Thermotoga* species have been well-documented. Although a robust genetic system for *Thermotoga* species has not been developed (Noll and Vargas 1997; Conners et al. 2006), the annotation of the *T. maritima* genome (Nelson et al. 1999) has allowed for significant functional genomic analysis of *T. maritima* with respect to metabolic regulation as well as other interesting phenomena such as the heat shock response, exopolysaccharide formation, biofilm formation, and quorum sensing activities (Chhabra et al. 2003; Nguyen et al. 2004; Pysz et al. 2004; Pysz et al. 2004; Conners et al. 2005; Nanavati et al. 2005) .

Mathematical modeling of H_2 production processes is limited. Unstructured kinetic and statistical modeling approaches have been applied, but the majority of these studies are on mixed cultures (Lay 2000; Chen et al. 2001; Lay 2001; Lay et al. 2005; Van Ginkel and Logan 2005; Chen et al. 2006; Mu et al. 2006; Whang et al. 2006). Metabolic models for the purpose of modeling H_2 production have only

recently been of interest (Manish et al. 2007; Oh et al. 2008; Zhang et al. 2009) and no unstructured or structured mathematical model has been developed for H_2 production by *T. neapolitana*.

In silico modeling of metabolism based on constraint-based analysis can be conducted for *T. neapolitana*. This requires reconstruction of the metabolic network and representation of this network as a system of linear equations defined by the stoichiometric matrix and a flux vector representing the rates associated with the network reactions. The reconstructed model can be converted into a computational model by the application of constraints. An optimal solution to the computational model can then be obtained by using linear programming to solve a particular objective function. This constraint-based analysis strategy has the major advantage of requiring no knowledge of the kinetic parameters for the network reactions, which are difficult to obtain (Teusink et al. 2000; Edwards et al. 2002; Patil et al. 2004; Heijnen 2005). Additional thermodynamic and regulatory constraints can be added to improve the predictive power of the model (Covert et al. 2001; Herrgård et al. 2004; Patil et al. 2004; Price et al. 2004; Feist et al. 2007; Henry et al. 2007). High-throughput data analysis methods can also be used to iteratively improve these models, including transcriptomics and proteomics for the provision of regulatory information (1999; Dharmadi and Gonzalez 2004) and fluxomic analysis to improve the flux constraints in the model (Sauer 2004).

Chapter 3

The Fermentation Stoichiometry of *Thermotoga neapolitana* and Influence of Temperature, Oxygen, and pH on Hydrogen Production[1,2]

ABSTRACT

The hyperthermophilic bacterium, *Thermotoga neapolitana*, has potential for use in biological hydrogen (H_2) production. The objectives of this study were to (1) determine the fermentation stoichiometry of *T. neapolitana* and examine H_2 production at various growth temperatures, (2) investigate the effect of oxygen (O_2) on H_2 production, and (3) determine the cause of glucose consumption inhibition. Batch fermentation experiments were conducted at temperatures of 60, 65, 70, 77, and 85 °C to determine product yield coefficients and volumetric productivity rates. Yield coefficients did not show significant changes with respect to growth temperature and the rate of H_2 production reached maximum levels in both the 77 and 85 °C experiments. The fermentation stoichiometry for *T. neapolitana* at 85 °C was 3.8 mol H_2, 2 mol CO_2, 1.8 mol acetate, and 0.1 mol lactate produced per mol of glucose consumed. Under microaerobic conditions H_2 production did not increase when compared to anaerobic conditions, which supports other evidence in the literature that *T. neapolitana* does not produce H_2 through microaerobic metabolism. Glucose consumption was inhibited by a decrease in pH. When pH was adjusted with buffer addition cultures completely consumed available glucose.

[1] Reproduced with permission from Biotechnology Progress, accepted for publication January 2009

[2] See Appendix 1 for revisions

3.1 Introduction

Large-scale hydrogen production methods include steam reforming of methane in natural gas and gasification of coal (Turner 2004). However both of these technologies use nonrenewable resources, thus there is considerable interest in the development of sustainable hydrogen (H_2) production methods based on renewable resources. A multitude of bacteria have the capability to produce H_2 from sugar substrates or biomass wastes (Hallenbeck and Benemann 2002; Hawkes et al. 2002; Nath and Das 2004; Hallenbeck 2005). The primary metric for an optimal H_2 producer is the H_2 yield coefficient, which is generally defined as the moles of H_2 produced per mole of substrate consumed. The model substrate that is used for comparison is glucose and more complex sugars are usually converted into glucose units for comparison between studies. The theoretical maximum H_2 yield coefficient, 4 mol H_2/mol glucose, is obtained when acetate is the primary volatile fatty acid (VFA) end-product of fermentation according to the following reaction:

$$C_6H_{12}O_6 + 4H_2O \rightarrow 2CH_3COO^- + 2HCO_3^- + 4H^+ + 4H_2 \qquad (3.1)$$

The $\Delta G^{\circ\prime}$ of this reaction is -206.3 kJ·mol^{-1} (Thauer et al. 1977). Few species approach this limit due to thermodynamic limitations and metabolic requirements such as maintaining a supply of NAD^+, which can be met by lactate production. Hallenbeck (2005) analyzed the reported hydrogen yield coefficients for several species and concluded that overall the yield coefficients for thermophilic microorganisms were closer to the thermodynamic limit than mesophilic species, where yields were closer to 2 mol H_2 per mol glucose. Additional advantages to using thermophilic organisms for hydrogen production are that at elevated temperatures there is reduced likelihood of contamination by mesophilic organisms (Claassen et al.

1999) and reduced affinity for the uptake of molecular H_2 by hydrogenases(Adams 1990; Claassen et al. 1999).

One of the thermophilic organisms under consideration for biological hydrogen production systems is *Thermotoga neapolitana* (Van Ooteghem 2005). *T. neapolitana* is a rod-shaped, Gram-negative bacterium that grows between 55 and 90 °C (Belkin et al. 1986; Jannasch et al. 1988). *T. neapolitana* consumes a variety of carbohydrate substrates including ribose, xylose, glucose, sucrose, maltose, lactose, galactose, starch, and glycogen (Jannasch et al. 1988). Although some of the metabolic products of *T. neapolitana* have been analyzed (Ravot et al. 1996; Van Ooteghem et al. 2002; Van Ooteghem et al. 2004; Eriksen et al. 2008), complete carbon and electron equivalents balances are not currently available for this species and are presented here. Such data is available for the related organism *T. maritima* which has been analyzed extensively, including its fermentation products (Schröder et al. 1994), metabolic pathways (Selig et al. 1997), and its annotated genome sequence (Nelson et al. 1999; Nelson et al. 2001; Conners et al. 2006). The metabolism of *T. neapolitana* has been under less scrutiny and if it is to be developed as a practical organism for H_2 production as proposed (Van Ooteghem 2005), then key fermentation parameters such as the H_2 yield coefficient and volumetric productivity rate, and a balanced fermentation stoichiometry need to be established. We chose to evaluate these parameters by conducting batch fermentation experiments at temperatures across the permissive growth range, because it cannot always be assumed that the optimum temperatures for growth and product formation are the same (Wang et al. 1979). In studies focusing on growth of *T. neapolitana*, where fermentation products were not quantified, the optimal temperature for growth was reported to be 77 °C (Belkin et al. 1986; Childers et al. 1992), whereas in studies focused on the H_2 production capability

of *T. neapolitana* 70 °C was used (Van Ooteghem et al. 2002; Van Ooteghem et al. 2004).

O_2 requirements are another issue to consider for biological hydrogen production. Strict anaerobes such as *Clostridia* sp. require that no oxygen is present. Prevention of O_2 exposure could be difficult on an industrial scale and requires expensive reducing agents (Kapdan and Kargi 2006); however potential H_2 yields in strict anaerobes are considerably higher than yields in facultative anaerobes that are O_2 tolerant, such as *Enterobacter* sp. Microaerobic species, that require O_2 partial pressures below atmospheric levels for aerobic metabolism, have been described such as *Klebisalle oxytoca* HP1 (Minnan et al. 2005). The question of oxygen tolerance and potential microaerobic metabolism of *T. neapolitana* and *T. maritima* has been raised in several studies (Belkin et al. 1986; Childers et al. 1992; Van Ooteghem et al. 2002; Eriksen et al. 2008; Le Fourn et al. 2008; Tosatto et al. 2008) and is addressed here. Belkin et al. (1986) and Childers et al. (1992) found that it was possible to maintain cultures under aerobic conditions at temperatures below the temperature range for growth. Belkin et al. (1986) found that cell death occurred under aerobic conditions at temperatures in the permissible growth range. Recent studies also suggest that O_2 tolerance is a characteristic of both *T. neapolitana* (Tosatto et al. 2008) and *T. maritima* (Le Fourn et al. 2008). The separate issue of possible microaerobic metabolism was raised by Van Ooteghem et al. (2002). They reported that the headspace O_2 concentration decreased during *T. neapolitana* growth and $Y_{H_2/Glucose}$ for these conditions was significantly higher than the theoretical 4 mol H_2/mol glucose that is possible from fermentative metabolism. Therefore, Van Ooteghem et al. (2002) hypothesized that the higher yield resulted from an unidentified aerobic pathway and suggested that *T. neapolitana* performs microaerobic metabolism. Experiments on the

effect of changing the headspace O_2 concentration during growth were conducted to test this hypothesis.

Another crucial consideration for economical industrial fermentations is achieving maximum substrate utilization. Limitations to glucose consumption were observed during the temperature experiments and have also been observed elsewhere (Schröder et al. 1994; Eriksen et al. 2008). It has been suggested that the limitation was due to pH, but no experimental evidence was provided (Eriksen et al. 2008). To verify this hypothesis we tested for inhibitory effects of both glucose and pH on glucose consumption.

3.2 Materials and Methods

3.2.1 Microorganism and Culture Medium

Thermotoga neapolitana DSM 4359 was obtained from the DSMZ culture collection and grown under anaerobic conditions in 160 mL serum bottles filled with 50 mL of *Thermotoga* basal medium with yeast extract and glucose (TBYG) (Childers et al. 1992). The growth medium was modified from the original formulation with an initial glucose concentration of 14 mM (2.5 $g \cdot L^{-1}$) rather than 28 mM (5 $g \cdot L^{-1}$). Unless otherwise noted, the initial glucose concentration for all experiments was 14 mM glucose. The growth medium was titrated with NaOH to obtain an initial pH of 7.5 for all experiments.

Culture medium was bubbled with nitrogen (N_2) for 15 min in the serum bottles to create anaerobic atmospheres. The bottles were then sealed with butyl rubber stoppers and aluminum crimp caps and autoclaved. In the present study sterile anoxic yeast extract (10% w/v) was autoclaved separately from other medium

components to avoid high temperature interactions during autoclaving between the glucose and peptides. Prior to inoculation, the yeast extract stock solution was added to all cultures for a concentration of 0.5 $g \cdot L^{-1}$ yeast extract.

3.2.2 Temperature Effects and Fermentation Balance

Batch growth experiments were performed to assess the influence of temperature on the H_2 yield of *T. neapolitana* and to obtain carbon balance data. Experiments were done at temperatures of 60, 65, 70, 77, and 85 °C on cultures grown in TBYG media. A set of control cultures were also grown on medium without glucose (TBY); the results from TBY cultures were subtracted from the results of TBYG cultures at each time point. This correction was made to account for the ability of the bacteria to metabolize yeast extract, an undefined medium component, which should not be included in carbon balance calculations. Replicate culture bottles were sacrificed for analysis, at eight time points for a 60 h period. In the 85 °C experiments four replicate cultures were analyzed for each time point (n = 4). For the 77 °C experiments n = 3 and n = 2 for the 60, 65, and 70 °C experiments.

For each set of temperature experiments frozen culture stocks were used to inoculate preliminary cultures in three bottles, with a 1% v/v inoculum. Bottle caps were flame-sterilized and sterile tuberculin syringes were flushed repeatedly with sterile N_2 before inoculation. After inoculation the preliminary cultures were incubated for 24 h at 77 °C in an oven (Barnstead Thermolyne, Dubuque, IA) for all experiments to achieve a high-density metabolically-active cell population. After 24 h of growth one of the preliminary cultures was selected to inoculate the experimental cultures. An orbital shaking water bath (Gyrotory Shaker, New Brunswick Scientific, Edison,

NJ) set at ~75 rpm and the temperature under study was used to mix and incubate the reactor contents.

3.2.3 Oxygen Exposure Effects

To determine the effects of oxygen exposure on *T. neapolitana*, four static batch cultures were incubated at 77 °C in an oven. Medium preparation, inoculation procedure, and culture liquid and headspace volumes were identical to those in the temperature experiments. In the O_2 exposure experiments the experimental cultures were incubated at 77 °C without agitation.

After 12 h of growth the headspace in each culture was sampled for H_2. Following this measurement, headspace volume (32 mL) was removed from all cultures using a sterile N_2-flushed syringe. In two of the cultures the same volume of sterile N_2 was reintroduced to the headspace, in the other two cultures atmospheric air was added with a sterile syringe to create 6% O_2 headspaces (32 mL air). The cultures were then re-incubated for a total period of 120 h and the headspace was periodically sampled for H_2 measurements. Turbidity and protein assay measurements were performed on each of the four cultures in triplicate.

3.2.4 Inhibition of Glucose Consumption

In temperature effect and carbon balance experiments, the available glucose was not completely consumed. To determine if the glucose itself was limiting glucose consumption, batch fermentation experiments were performed with different initial glucose concentrations, 14, 11, 8.3, and 5.6 mM (corresponding to 2.5, 2.0, 1.5 and 1.0 $g \cdot L^{-1}$). Liquid phase 0.5 mL samples were removed over a period of 60 h to

quantify glucose consumption as a function of time. To determine if pH was limiting glucose consumption in batch cultures, cultures were grown with 14 mM glucose at 77 °C under 3 different sets of conditions. Control cultures contained 6 $g \cdot L^{-1}$ $PIPES \cdot 1.5Na$, 1.5x buffer cultures contained 9 $g \cdot L^{-1}$ $PIPES \cdot 1.5Na$, and sodium bicarbonate ($NaHCO_3$) injection cultures contained 6 $g \cdot L^{-1}$ $PIPES \cdot 1.5Na$ and received a 1 mL injection of sterile anoxic 1 M $NaHCO_3$ after 30 h of growth. Glucose and H_2 measurements were made using 0.5 mL liquid phase and headspace samples taken at several time points over a 90 h growth period. After 90 h of growth, pH was measured in all cultures to confirm that an adjustment was made in the 1.5x buffer and $NaHCO_3$ injection cultures. H_2 yield from glucose was calculated for 48 h, to determine if pH adjustment had any effect on H_2 yield. Both the initial glucose concentration and pH adjustment experiments cultures were replicated (n = 4 for each glucose concentration and for each pH condition) and grown at 77 °C at 75 rpm under anaerobic conditions in the orbital water bath as in the temperature effect experiments.

3.2.5 Analytical Methods

Total protein and turbidity measurements were used to quantify biomass production in temperature effects experiments. Turbidity was quantified by measuring the absorbance of triplicate 200 μL culture samples at 670 nm. A modified Bradford method (Stoscheck 1990), described here, was used to measure total protein concentration in the cell cultures. From each culture triplicate 2 mL samples were removed and placed in screw-cap centrifuge tubes. The samples were centrifuged at 13,000 g for 8 min and the supernatant was removed for substrate and co-product analysis. The cell pellet was resuspended in 0.5 mL of deionized H_2O and the centrifugation step was repeated. The washed cell pellet was resuspended again in

H_2O and 20 μL of this cell suspension was removed for the Bradford protein assay. Bovine serum albumin (1 $mg \cdot mL^{-1}$ from Sigma) was used as the standard. Samples and standards were treated with 50 μL of 1 M NaOH to solubilize the cell membranes in the samples. The Bradford dye reagent was added and the absorbance at 595 nm was measured using a Synergy HT microplate reader (Bio-tek Instruments, Winooski, VT).

Hydrogen production was measured with a gas chromatograph (model 310, SRI Instruments, Torrance, CA) equipped with a thermal conductivity detector and a 3 ft Molesieve 5A column. From each culture 0.5 mL headspace samples were removed using a gastight syringe (Hamilton or Henke Sass Wolf GmbH). The carrier gas was N_2 and the column oven and detector temperatures were 27 and 100 °C respectively. Calibration curves were generated using pure H_2 standards to estimate the percentage of the headspace gas that was H_2. To account for the increase in headspace pressure due to gas production, the total headspace pressure was measured using pressure transducers coupled to a data acquisition system (NI SCXI-1000, National Instruments, Austin, TX). The moles of H_2 produced were estimated using the H_2 percentage and pressure measurements and the ideal gas law.

Due to the high degree of solubility of CO_2 in water, total CO_2 was measured using a Total Organic Carbon (TOC) analyzer (Shimadzu TOC-V_{CPH}, Columbia, MD) and the following procedure based on the work of Huber et al. (1986). After all other analyses had been completed 0.1 mL of 4 M NaOH was injected into 60 h, 0 h, and uninoculated bottles to raise the pH above 8.3 and force all gas phase carbon dioxide into the liquid phase. After this pH adjustment, the bottles were shaken on ice and in a cold room at 150 rpm for 90 min. Triplicate liquid samples were taken from each culture and filtered through 0.2 μm polyethersulfone syringe filters. The culture samples were diluted 1:40 in 18 MΩ Millipore filtered water to reduce the NaCl

concentration to a level permissible for total inorganic carbon analysis using the TOC analyzer. Multiple measurements were made on each sample. A 1000 $mg \cdot L^{-1}$ inorganic carbon standard was prepared using sodium carbonate and sodium bicarbonate and manually diluted to 10 $mg \cdot L^{-1}$. The TOC analyzer performed subsequent automatic dilutions of the 10 $mg \cdot L^{-1}$ solution to generate a calibration curve for inorganic carbon. The CO_2 measurements include all inorganic carbon produced in the system, both gas and liquid phase CO_2 and carbonates.

The concentrations of glucose and organic acids in the cultures were quantified using a Shimadzu HPLC system (Shimadzu, Columbia, MD) equipped with a refractive index detector and a Bio-Rad Fast Acid Analysis column. The system was operated at 25 °C with a mobile phase of 5 mM H_2SO_4 at a flow rate of 0.6 $mL \cdot min^{-1}$. Two sets of standards were used, glucose in 5 mM H_2SO_4 and an acid mixture containing lactate, acetate, propionate, ethanol and butyrate. All standards and samples were filtered through 0.2 μm syringe filters prior to analysis. All substrate and product concentration results were expressed as millimoles per liter of liquid volume.

3.2.6 Calculation of Yield Coefficients, Volumetric Productivity, and Fermentation Balances

Yield coefficients were defined as the ratio of production of H_2, acetate, lactate, and biomass to consumption of glucose, $Y_{H_2/Glucose}$, $Y_{Acetate/Glucose}$, $Y_{Lactate/Glucose}$, and $Y_{X/Glucose}$. The yield coefficients were estimated using the product formation and glucose consumption rates, $r_{Product}$ and $r_{Glucose}$, from 0 h to 48 h. This method is illustrated by the equation

$$Y_{Product/Glucose} = \frac{r_{Product}}{r_{Glucose}} \tag{3.2}$$

where

$r_{Product}$ = rate of product formation ($mmol \cdot L^{-1} \cdot h^{-1}$) and

$r_{Glucose}$ = rate of glucose consumption ($mmol \cdot L^{-1} \cdot h^{-1}$).

Another useful parameter for gauging the performance of batch fermentation is volumetric productivity (Wang et al. 1979). The overall volumetric productivity, $Q_{Product}$, includes the full time period of the fermentation from the lag phase until termination and is defined by

$$Q_{Product} = \frac{m_{Product}}{V \cdot t} \tag{3.3}$$

where

$m_{Product}$ = mass of product (mmol),

V = liquid volume (L), and

t = total time of fermentation (h).

This metric was calculated for each product based on the endpoint of 48 h. The carbon, and electron equivalent balances for 60 h cultures were calculated based on the methods (Gottschalk 1986; Rittmann and McCarty 2001) that are briefly described here. The carbon balance is the ratio of the moles of carbon recovered from the fermentation products to the moles of carbon present in the substrate. The moles of carbon for a compound are calculated by multiplying the number of carbon molecules in the molecular formula by the number of moles of the compound that were experimentally measured. An electron equivalents balance is the ratio of electron equivalents in the fermentation products to the electron equivalents in the substrate. The electron equivalents per mole for a compound are calculated by oxidizing that compound to CO_2. For example, when 1 mole of glucose is oxidized to CO_2, 24 e^- are released in the reaction; therefore glucose has 24 electron equivalents per mole, the electron equivalents are then determined by multiplying this value by the

measured moles of glucose. The mmol of biomass was calculated from the mg cellular protein measured in the Bradford protocol, based on the assumptions that dry cell weight of the bacteria is ~50% protein (Gottschalk 1986) and that the chemical composition of the bacteria can be approximated by the empirical formula $C_5H_8O_3NS_{0.05}$ based on elemental analysis of *T. maritima* (Rinker and Kelly 2000).

3.3 Results

3.3.1 Temperature Effects on Batch Growth

Biomass production in the batch cultures was indirectly observed through temporal measurements of protein and turbidity as plotted in Figure 3.1A and 3.1B, respectively. The two data sets show negligible biomass accumulation in the 60 °C cultures over the 60 h growth period, and rapid growth, reaching stationary phase at 24 h, in the 77 and 85 °C cultures. Flocculation of the bacteria resulted in non-uniform cell samples and made accurate biomass measurements challenging.

Glucose concentrations measured at different time points and temperatures are plotted in Figure 3.2A. Measurements were more consistent than those obtained for protein (Figure 3.1A) and turbidity (Figure 3.1B). Glucose utilization was observed at all temperatures, with the 77 and 85 °C cultures exhibiting the greatest rate and extent of glucose utilization. The term extent refers to the extent of conversion of reactants to products as used in the field of chemical engineering. The HPLC results indicated that glucose consumption was incomplete in the 77 and 85 °C cultures. The glucose concentrations measured using the HPLC method were similar to those obtained with the colorimetric p-hydroxybenzoic acid hydrazide (PAHBAH) assay (Lever 1972) (data not shown). In addition, cultures with a 28 mM (5 $g \cdot L^{-1}$) initial glucose

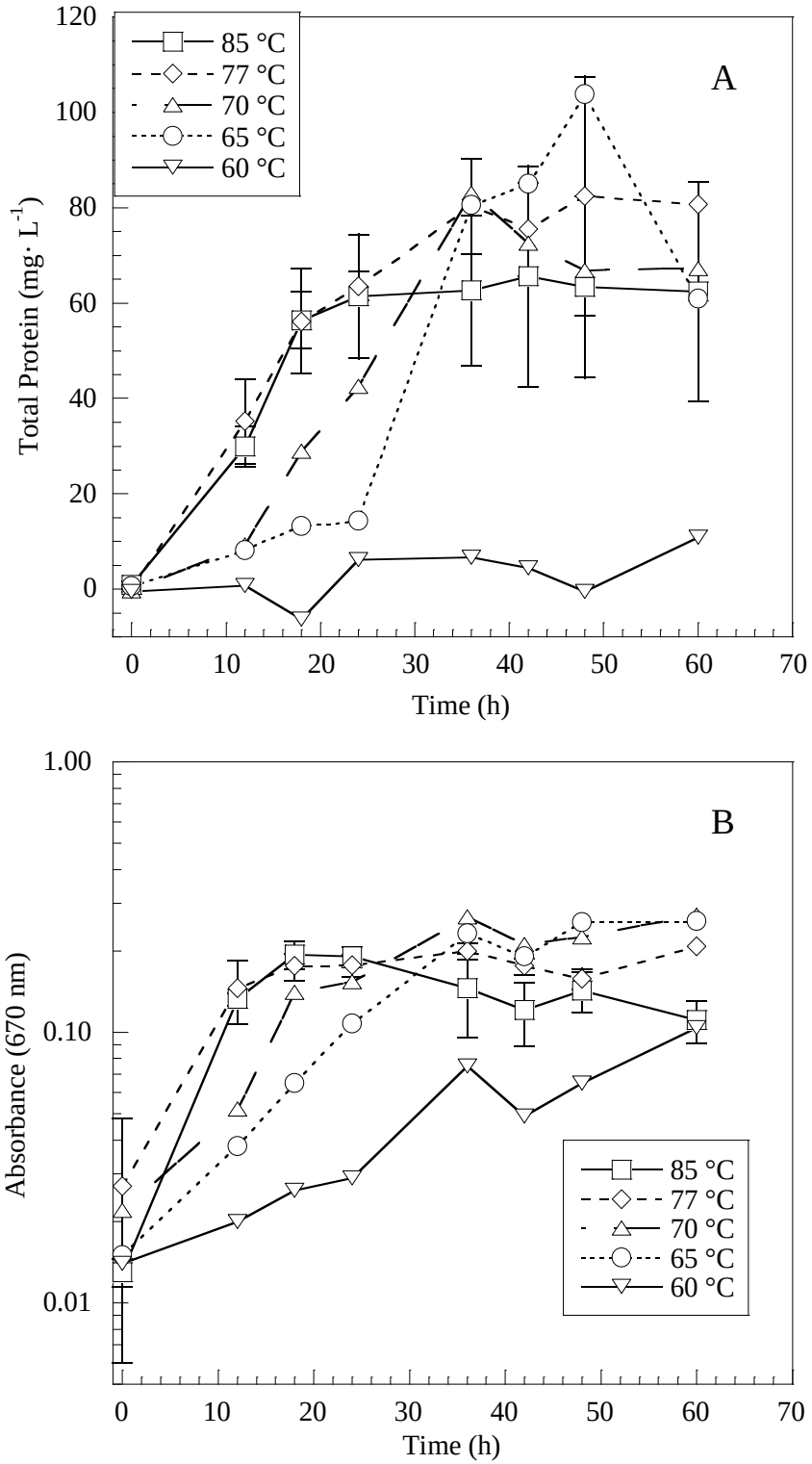


Figure 3.1 Protein concentration (A) and Absorbance (670 nm) (B) measurements as a function of time for different temperatures. Error bars represent standard deviation of replicate cultures (n = 3 for 77 °C and n = 4 for 85 °C).

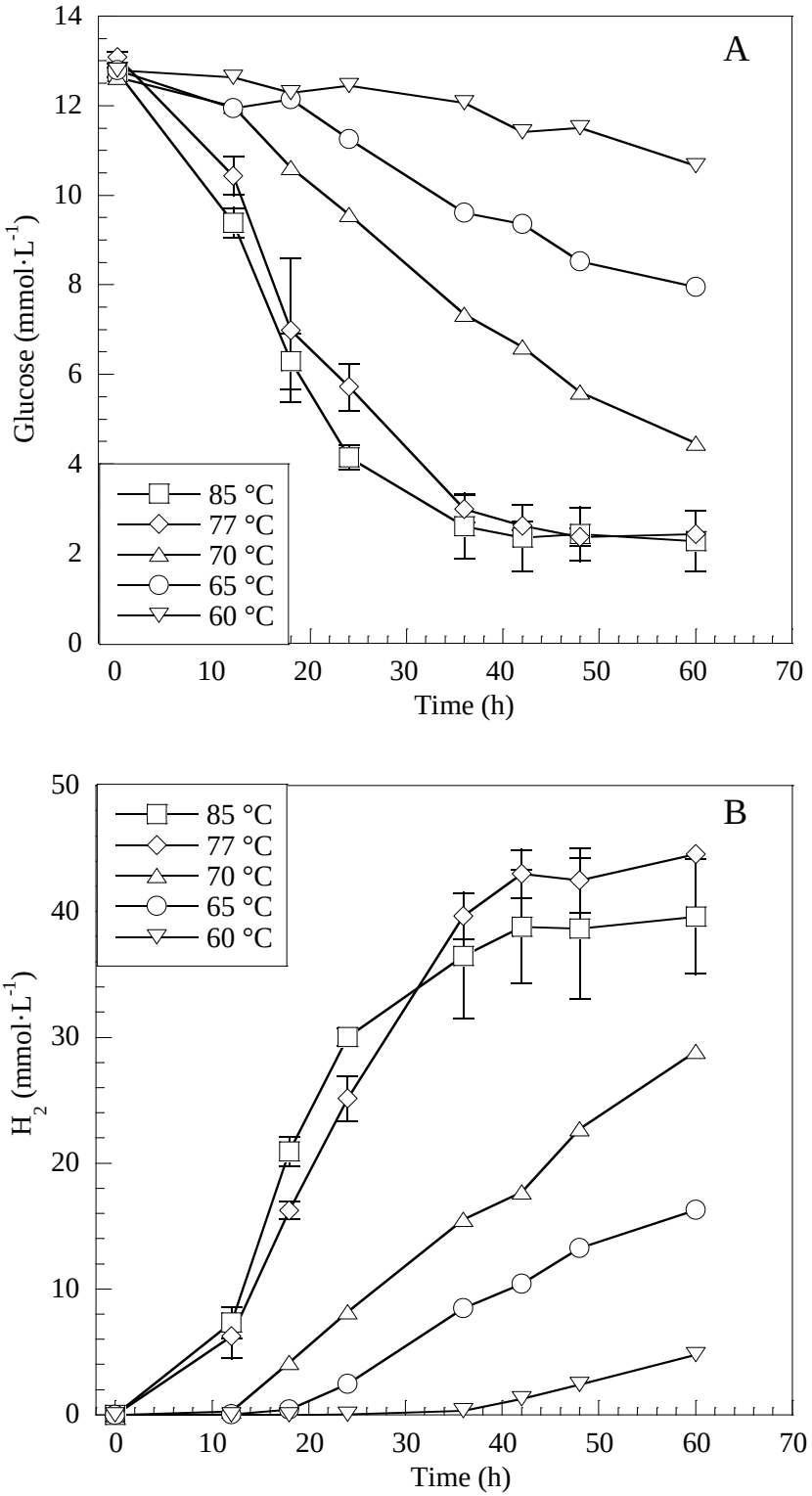


Figure 3.2 Glucose concentration (A) and H_2 concentration (B) measurements as a function of time for different temperatures. Error bars represent standard deviation of replicate cultures (n = 3 for 77 °C and n = 4 for 85 °C).

concentration consumed similar quantities of glucose (data not shown). Both these results suggest that the incomplete degradation is not due to detection limits of the HPLC system.

Temporal H_2 measurements at different temperatures are plotted in Figure 3.2B. Cultures grown at 77 and 85 °C exhibited the greatest rate and extent of H_2 production reaching the maximum H_2 concentration by 42 h.

In Figure 3.3A and 3.3B temporal concentration measurements are plotted for acetate and lactate, respectively. Although the rate and extent of glucose utilization and H_2 production at 77 and 85 °C were nearly identical, the rate and extent of organic acid production was different for the two temperatures. The 85 °C cultures exhibited a higher rate and extent of production.

3.3.2 Yield Coefficients and Volumetric Productivity

The yield coefficients and the volumetric productivities are listed in Table 3.1. The standard deviations are presented for all temperatures including the duplicate cultures so that variability between temperatures can be assessed. Due to the negligible biomass accumulation and extended lag period in the 60 °C cultures these values were eliminated from further analysis. The yield coefficients did not show a significant change with respect to growth temperature and the volumetric productivities for hydrogen, acetate, and lactate increased as growth temperature increased.

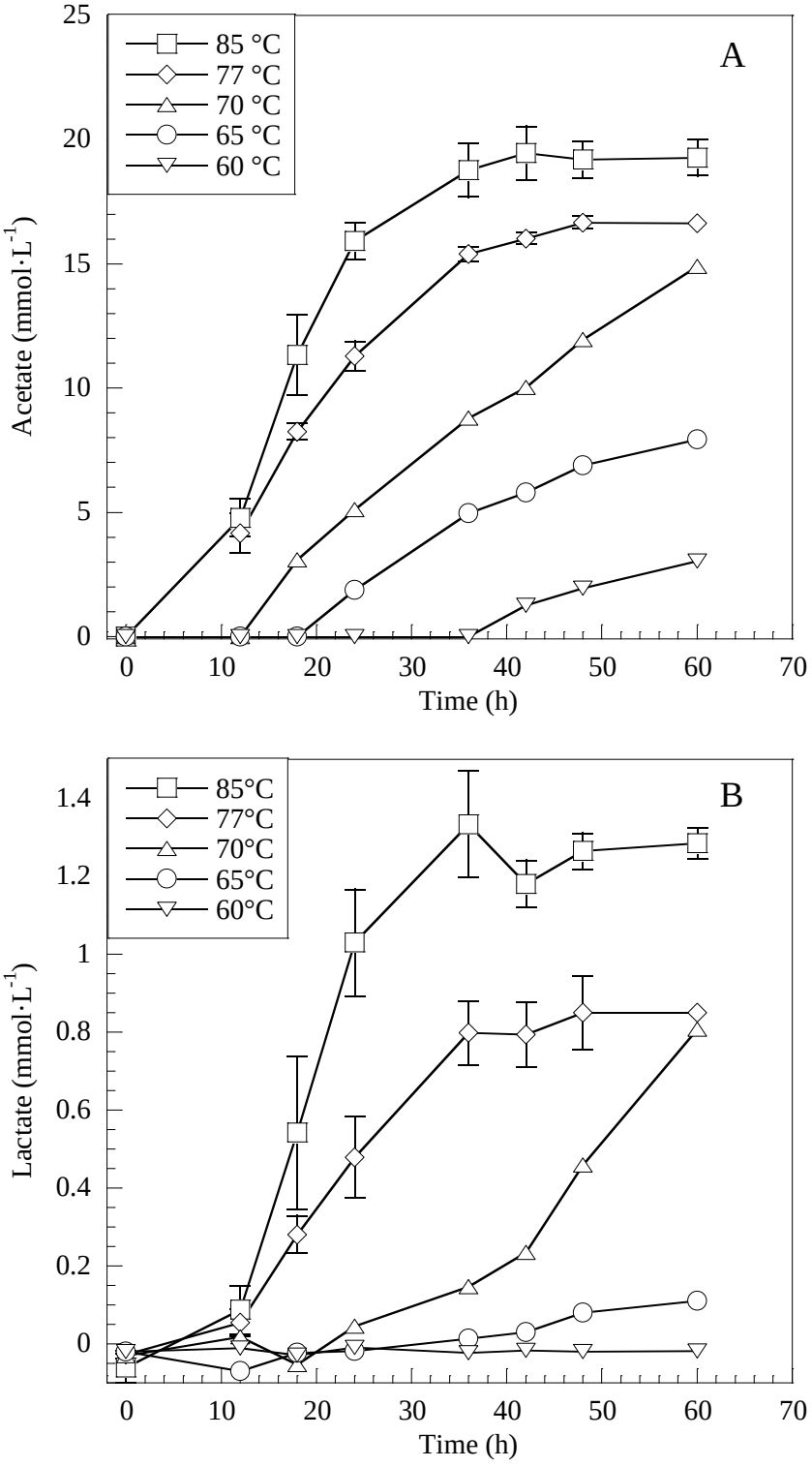


Figure 3.3 Acetate (A) and (B) lactate concentration measurements as a function of time for different temperatures. Error bars represent standard deviation of replicate cultures (n = 3 for 77 °C and n = 4 for 85 °C).

Table 3.1 Yield coefficients of products per mol of glucose consumed and volumetric productivity at various growth temperatures after 48 h of growth. Standard deviations represent culture replicates.

	Temperature				
	60 °C	**65 °C**	**70 °C**	**77 °C**	**85 °C**
Yield Coefficients (mol/ mol)					
$Y_{H_2/Glucose}$	2.04 ± 0.05	3.09 ± 0.30	3.18 ± 0.02	3.85 ± 0.28	3.75 ± 0.49
$Y_{Acetate/Glucose}$	2.57 ± 0.52	1.62 ± 0.18	1.70 ± 0.06	1.56 ± 0.05	1.87 ± 0.11
$Y_{Lactate/Glucose}$	n.d.[a]	0.019 ± 0.004	0.06 ± 0.02	0.08 ± 0.01	0.12 ± 0.01
$Y_{X/Glucose}$	n.d.	0.37 ± 0.027	0.07 ± 0.037	0.12 ± 0.002	0.09 ± 0.030
Volumetric Productivity ($mmol \cdot L^{-1} \cdot h^{-1}$)					
Q_{H_2}	0.05 ± 0.009	0.27 ± 0.003	0.47 ± 0.01	0.86 ± 0.04	0.80 ± 0.10
$Q_{Acetate}$	0.06 ±0.004	0.144 ± 0.0003	0.249 ± 0.0004	0.347 ± 0.005	0.40 ± 0.02
$Q_{Lactate}$	n.d.	0.002 ± 0.0001	0.010 ± 0.004	0.018 ± 0.002	0.026 ± 0.001
Q_X	n.d.	0.008 ± 0.0006	0.002 ± 0.00003	0.002 ± 0.0008	0.002 ± 0.0006

[a] not detected

3.3.3 Fermentation Balance

The carbon and electron equivalents balances were determined at different growth temperatures for *T. neapolitana*; results are shown in Table 3.2. The carbon balances were at acceptable levels for the 85, 70, and 65 °C cultures. The carbon balance for the 77 °C cultures was not shown, because CO_2 measurements were not taken for these experiments. The electron equivalent balances for all four temperatures were also at acceptable levels.

Table 3.2 Fermentation balance results for 60 h cultures.

Temperature (°C)	Products (mmol/mmol Glucose)					Carbon Balance	Electron Equivalents Balance
	Acetate	Lactate	Biomass[a]	H_2	CO_2		
85	1.84	0.12	0.09	3.76	2.35	1.14	1.06
77	1.57	0.08	0.12	4.11	n.d.[b]	n.d.	1.00
70	1.83	0.10	0.06	3.47	1.98	1.04	1.00
65	1.64	0.02	0.19	3.34	1.94	1.04	1.00

[a] mmol Biomass was calculated based on the empirical formula $C_5H_8O_3NS_{0.05}$ for *T. maritima*
[b] n.d - not determined

3.3.4 O_2 Exposure Effect

Hydrogen concentration measurements as a function of time for different O_2 exposure levels are plotted in Figure 3.4. All cultures continued producing H_2 following the headspace adjustment with either sterile air or N_2. The rates of accumulation and maximum quantity of H_2 were significantly different; the anaerobic cultures outperformed the O_2-exposed cultures in both categories. At the end of the

120 h incubation period the mean of the turbidity measurements from the 4 cultures had a 6% coefficient of variance (CV) and the protein concentration results had a 7% CV, indicating that O_2 exposure after 12 h of growth did not significantly decrease the quantity of cell mass that was accumulated (data not presented). Other fermentation products were not measured in these experiments.

3.3.5 Inhibition of Glucose Consumption by Low pH

Glucose concentration measurements from experiments with 14 mM initial glucose concentration indicated that glucose consumption was inhibited. The glucose itself was not the inhibiting factor, because experimentation with a range of initial glucose concentrations resulted in complete consumption of all available glucose in cultures with 8.3 and 5.6 mM initial glucose concentrations after 40 h of growth (data not shown). Inhibition from low pH was tested in pH adjustment experiments, the glucose and H_2 concentration results from these experiments are shown in Figure 3.5. Both methods of adjusting pH influenced the extent of glucose consumption and H_2 production. When pH was adjusted all the glucose was consumed by 68 h, and more H_2 was produced, therefore the yield did not vary significantly compared to the control cultures. After 90 h pH measurements were made to verify that adjustments were made. The pH for the control cultures was 5.12 ± 0.01, whereas the pH for the 1.5x buffer cultures was 5.54 ± 0.01 and for the $NaHCO_3$ injection cultures it was 6.33 ± 0.14.

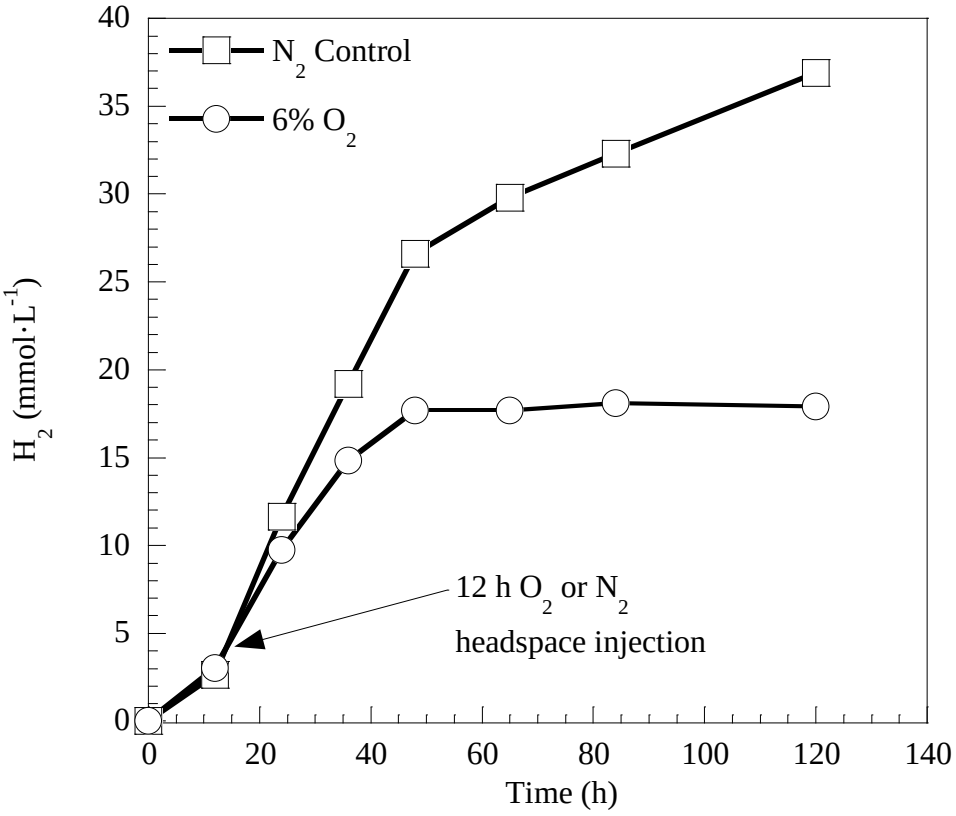


Figure 3.4 H_2 concentrations as a function of time for 77 °C cultures with either O_2 or N_2 injections. Data points represent the average of duplicate cultures.

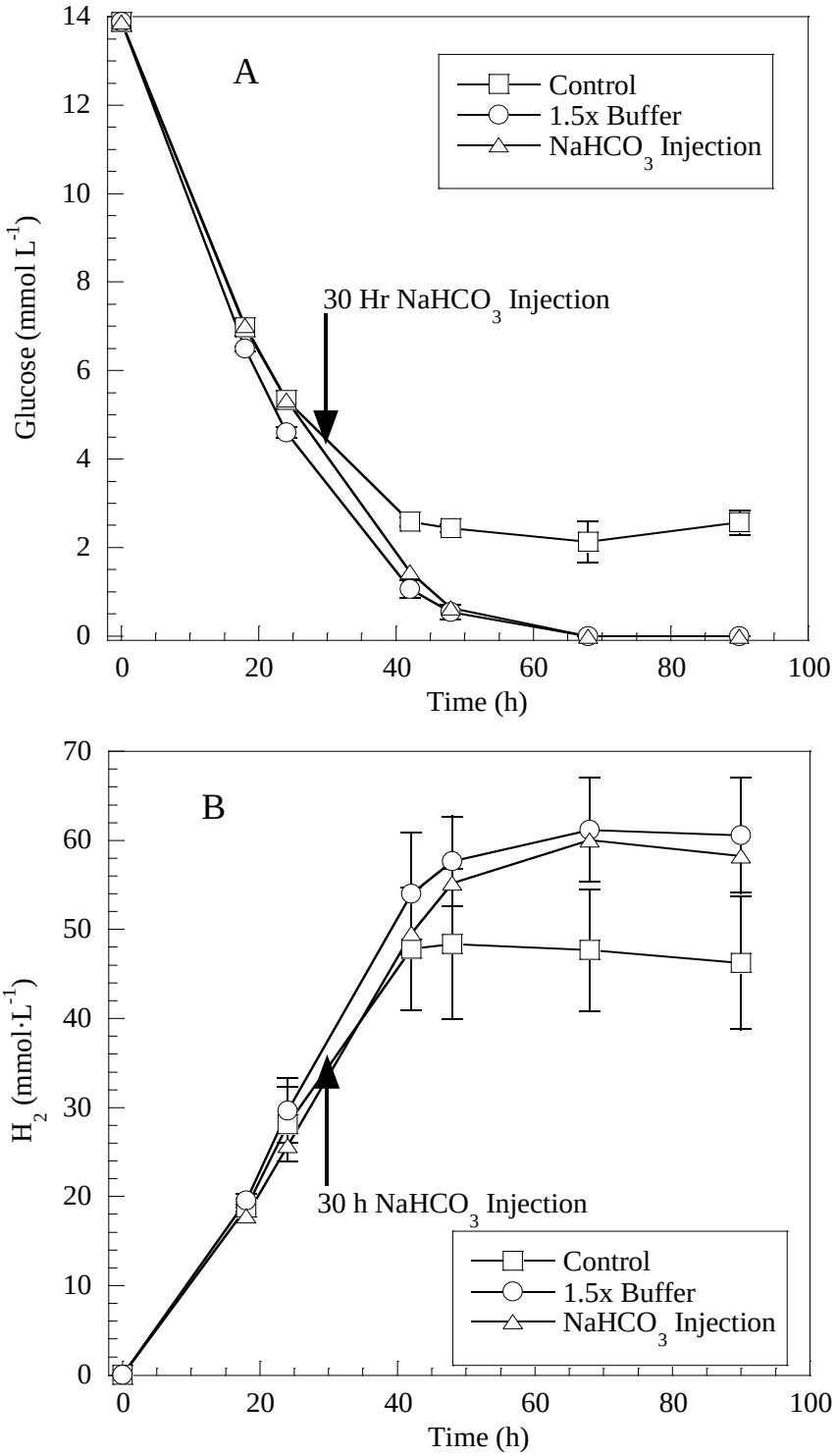


Figure 3.5 Glucose (A) and H_2 (B) concentration in control and pH adjusted cultures with either 1.5x initial buffer concentration or an injection of 1 mL 1M $NaHCO_3$ after 30 h of growth. Error bars represent standard deviation of replicate cultures (n = 4 for each condition).

3.4 Discussion

Fermentation data has been presented for *T. neapolitana* across the permissive growth temperature range. The primary fermentation products were H_2, CO_2, acetate, and small amounts of lactate. The rate and extent of glucose consumption and H_2 formation increased with increased operating temperature from 60 to 77 °C, but there was no significant difference between the results for the 77 and 85 °C cultures. In contrast to the H_2 production results, more acetate and lactate were found in the 85 °C cultures than in the 77 °C cultures. However, although the acetate and lactate concentration results indicated a difference between 77 and 85 °C (see Figure 3.3) a comparison of the $Y_{Acetate/Glucose}$ and $Y_{Lactate/Glucose}$ values for the operating temperatures between 65 and 85 °C (see Table 3.1) indicated that there was no significant change in acetate or lactate yields with operating temperature.

In previous studies the optimal temperature for growth of *T. neapolitana* was found to be 77 °C (Belkin et al. 1986; Childers et al. 1992), but there was no investigation into the effect of temperature on H_2 productivity in these studies. Researchers that focused on the H_2 production capability of *T. neapolitana* selected an operating temperature of 70 (Van Ooteghem et al. 2002; Van Ooteghem et al. 2004) or 80 °C (Eriksen et al. 2008), but did not quantify biomass formation and it was unclear why 70 °C was selected. A recent study on the influence of temperature on H_2 production of *T. neapolitana* and *T. maritima* (Nguyen et al. 2008) supports our result that H_2 production is linked with growth in this species. Our results indicated that there was not a significant change between volumetric productivity, Q_{H_2}, at 77 and 85 °C. Given the assumption that there would be increased heating costs to operate at temperatures above 77 °C and there would be no increase in production rate or yield,

operating at 77 °C is sufficient not only for maximum growth as shown by Childers et al. (1992), but also sufficient for maximum H_2 production rates.

Carbon and electron equivalent balance results were obtained to establish the fermentation stoichiometry of glucose for *T. neapolitana*. In previous studies some fermentation products of *T. neapolitana* were analyzed, but complete carbon and electron equivalents balances have not been presented (Ravot et al. 1996; Van Ooteghem et al. 2002; Van Ooteghem et al. 2004; Eriksen et al. 2008). The 85 °C fermentation balance of *T. neapolitana* showed that per mol glucose consumed, 1.8 mol acetate, 0.1 mol lactate, 3.8 mol H_2, and 2 mol CO_2 were produced. These results are consistent with those of Schröder et al. (1994), where 2 mol acetate, 2 mol CO_2, and 4 mol H_2 were formed per mol glucose in cultures of *T. maritima*. Eriksen et al. (2008) reported a H_2 yield of 2.4 ± 0.3 mol H_2/mol glucose *T. neapolitana*. This is lower than our result and that of Schröder et al. (1994).

The question of microaerobic metabolism in *T. neapolitana* has been raised in the literature with the observation of H_2 yields in this species that greatly exceed the theoretical limit for this anaerobic fermentation, and the hypothesis that this result is due to an unidentified aerobic pathway using O_2 as a terminal electron acceptor (Van Ooteghem et al. 2002). However, subsequent research (Eriksen et al. 2008) and the results of the present study contradict this hypothesis. The headspace composition experiments presented here confirm the observation of Eriksen et al. (2008) that H_2 production continues following the injection of O_2, but the rate and extent of H_2 production is lower than the results found in cultures without O_2. Therefore, *T. neapolitana* does not produce more hydrogen when O_2 is present and does not exceed the theoretical limits of fermentative H_2 production. A more realistic explanation for the observed decrease in O_2 headspace concentration is that the reducing agent used in the growth medium, cysteine, reduces the O_2 in the headspace.

This was shown by Eriksen et al. (2008), O_2 headspace concentrations decreased in abiotic serum bottles when cysteine was provided. In the O_2 amended cultures we observed a surface layer where the oxidized resazurin dye accumulated. This was also previously observed (Eriksen et al. 2008) and suggests that an anaerobic zone supporting *T. neapolitana* growth was present below the surface layer.

Incomplete glucose consumption was observed for the 77 and 85 °C cultures and was investigated further to determine the cause of inhibition. Experiments with different initial glucose concentrations confirmed that *T. neapolitana* could consume glucose completely, so it was not a property of the glucose itself that limited consumption. Experiments with pH adjustment showed that when cultures were neutralized with either increased initial buffer or injection of $NaHCO_3$, glucose was completely consumed and H_2 production increased proportionally; thus there was no change in H_2 yield from glucose in neutralized cultures when compared to acidic control cultures. This result suggests that pH control is a significant factor in maintaining *T. neapolitana* cultures that can consume available substrate.

3.5 Conclusion

Further insight has been offered regarding the H_2 production capabilities and the fermentation pathway of *T. neapolitana*. Yield coefficients for the fermentation products did not vary significantly with temperature and the maximum volumetric productivity rates were achieved at 77 and 85 °C, thus it was concluded that 77 °C was sufficient for maximum H_2 production. This result confirmed the link between cell growth and hydrogen production, because 77 °C is also optimal for growth (Belkin et al. 1986; Childers et al. 1992). A closed carbon balance had not been previously shown for this organism and was established based on results from multiple

temperature experiments. Complete recovery was obtained using the model substrate glucose and the relatedness of *T. neapolitana* to *T. maritima* was underscored by the similarity in fermentation stoichiometry. The effect of O_2 on H_2 production was also investigated and provided supporting evidence to contradict the hypothesis that *T. neapolitana* produces more H_2 under microaerobic conditions than in anoxic conditions. Incomplete glucose consumption was due decreased culture pH, when pH was increased, glucose was completely consumed. Future work should focus on continuous process control of pH and product removal with the goal of achieving maximum substrate utilization and H_2 production.

Chapter 4

Comparative Constraint-Based Model Development and Hypothesis Generation for *Thermotoga neapolitana*[3]

ABSTRACT

A comparative constraint-based model development method is presented; the constraint-based model for one organism is modified to create a model for a second organism based on synteny between the two annotated genome sequences. With this method a model was created for *Thermotoga neapolitana* using the model for *T. maritima* as a template. Twenty-five of the 479 genes in the *T. maritima* model did not have orthologs in *T. neapolitana* and the reactions associated with these genes were removed from the model. The *T. neapolitana* model was verified with flux balance analysis (FBA) and dynamic FBA. Sixteen carbon sources that supported FBA growth simulations for the *T. maritima* model did not support growth for the *T. neapolitana* model. This result indicates metabolic differences between the species. Dynamic FBA with the *T. neapolitana* model was in agreement with *T. neapolitana* batch growth experimental data. Metabolic hypotheses were developed with the *T. neapolitana* model. To reflect experimental evidence, the *T. neapolitana* model was amended to support growth with cysteine as a sulfur source. This capability is not present in *T. maritima*. The two species do have identical hydrogenase-related genes. Reactions for a bifurcating hydrogenase and a membrane bound NADH:Fd oxidoreductase were tested in the *T. neapolitana* model Similar to comparative

[3] Submitted for publication in Biotechnology and Bioengineering

genomics, the comparative model development approach is an expeditious method to generate constraint-based models for organisms based on synteny between species.

4.1 Introduction

Systems-level analysis of biological interaction networks, systems biology, offers a structured approach to elucidating metabolic networks and for the rational genetic engineering of important microbial systems for industrial biotechnology. One area where systems-level analysis is being deployed is in understanding metabolic pathways for biological hydrogen (H_2) production by hyperthermophilic microorganisms. Hyperthermophilic H_2 producers such as *Thermotoga maritima* and *Thermotoga neapolitana* are of particular interest due to the high H_2 yields obtained in these species, which approach the theoretical maximum of 4 mol H_2/ mol glucose (Thauer et al. 1977). Since the publication of the *Thermotoga maritima* genome (Nelson et al. 1999), numerous biochemical and functional genomics studies have been conducted with this *Thermotoga* type species and have been summarized in the excellent review by Conners et al. (2006). The genome annotation and extensive body of literature for *T. maritima* have been used for reconstruction of a detailed model of *T. maritima* central carbon metabolism (Zhang et al. 2009). Such a model serves as a logical contextual framework to store and probe biological data (Feist et al. 2007) and is advantageous not only to understand individual species, but can also be used to explore the similarities and differences between related organisms. For example, *in silico* models of different strains of *Staphylococcus aureus* were developed and compared to identify potential anti-microbial drug targets (2009). In the present study, the constraint based reconstruction and analysis (COBRA) modeling framework (Becker et al. 2007) is utilized to compare *T. maritima* and *T. neapolitana.*

These two *Thermotoga* species are highly similar. DeBoy et al. (2006) performed a whole genome alignment between *T. maritima* and *T. neapolitana* and found that 1,726 proteins are shared between the two strains, out of 1,838 total predicted proteins for *T. maritima* and 1,903 predicted proteins for *T. neapolitana*. The two proteomes have an average percentage of identity of 83.6% and an average percentage of similarity of 92.4% (DeBoy et al. 2006). Despite the high degree of sequence similarity, DeBoy and colleagues identified large-scale inversion events in the genomes (2006). In addition to genome rearrangement, key physiological differences between the species have been identified. For example, *T. neapolitana* grows on a defined medium with cysteine as a sulfur source (Childers et al. 1992), whereas *T. maritima* requires elemental sulfur for growth on defined medium both *in vivo* (Schröder et al. 1994) and *in silico* (Zhang et al. 2009). Elemental sulfur dependence is a drawback for use of *T. maritima* as a H_2 production organism because the presence of elemental sulfur leads to the production of the toxic corrosive gas H_2S (Schröder et al. 1994) at the expense of H_2 production (Huber et al. 1986; Käslin et al. 1998). The metabolic modeling approach presented here is used to theorize how *T. neapolitana* can directly metabolize cysteine.

The draft genome annotation for *T. neapolitana* has recently become publicly available (Lim et al. 2009), making it feasible to use the genome annotation and extensive literature searching for reconstruction of the *T. neapolitana* metabolic network. However, given that the reconstruction process is not a trivial endeavor (Thiele and Palsson 2010), we instead exploited the high degree of sequence similarity between the *T. neapolitana* and *T. maritima* species to develop the comparative model reconstruction approach presented in this paper. A key component in the comparative reconstruction approach is the use of the multiple genome alignment program, Mauve, which can handle large-scale sequence inversions to predict similarity between

genomes (Darling et al. 2004). Using Mauve allows the prediction of instances of synteny for sequence rearrangements that are not easily obtained in database searches and also allows the direct comparison of nucleotide sequences including the unannotated regions of genomes. The *T. neapolitana* model we created using the comparative modeling approach was used to test carbon substrate utilization in *T. neapolitana* and to generate hypotheses for potential pathways for cysteine metabolism and H_2 production.

4.2 Methods

4.2.1 Required Data and Software

GenBank genome annotation files for *T. maritima* and *T. neapolitana* were downloaded from the Integrated Microbial Genome (IMG) database hosted by the Joint Genomic Institute (JGI) (Markowitz et al. 2010). For analysis of the genome annotations the Mauve multiple genome alignment program (Darling et al. 2004) was used. The *T. maritima* central metabolism model was obtained from the supplementary info of Zhang et al. (2009). Matlab version 7.8 (The Mathworks Inc., Natick, MA) and the following freely available toolboxes were installed on a standard personal computer with the Windows 7 OS: the Constraint-based Reconstruction and Analysis (COBRA) toolbox (Becker et al. 2007), the SBML (Systems Biology Markup Language) toolbox (Keating et al. 2006), and the GLPK (Gnu Linear Programming Kit) package (Makhorin 2008). Instructions for installation are provided by Becker et al. (2007).

4.2.2 Comparative Model Reconstruction

The comparative model reconstruction pipeline is presented in Figure 4.1. The SBML model of *T. maritima* central metabolism was imported into the Matlab workspace using the COBRA toolbox. The Genbank genome annotation files for *T. maritima* and *T. neapolitana* were imported into the Mauve program and analyzed using the progressive alignment algorithm. An ortholog file was generated containing all genes in both genomes with the orthologous genes matched in the file. Software was developed in Matlab to import and process the Mauve ortholog file and to replace the genes in the *T. maritima* COBRA model with the orthologs from *T. neapolitana*.

The *T. maritima* genes that were not found to have synteny with *T. neapolitana* genes based on the Mauve ortholog report remained in the model and the effects of deleting each gene were tested. It was assumed for these tests that biomass formation was the objective function and glucose was provided as the carbon source with minimal growth medium. Genes that were essential and nonessential for growth were identified from this process. Both sets were manually searched in the Mauve program for synteny with *T. neapolitana* genes. *T. neapolitana* orthologs for all essential genes and most non-essential genes were identified in the Mauve alignment. *T. neapolitana* orthologs were added into the model by appending the *T. neapolitana* locus tag to the *T. maritima* locus tag. The remaining *T. maritima* genes that did not have orthologs in *T. neapolitana* and the corresponding reactions were not included in the new *T. neapolitana* model.

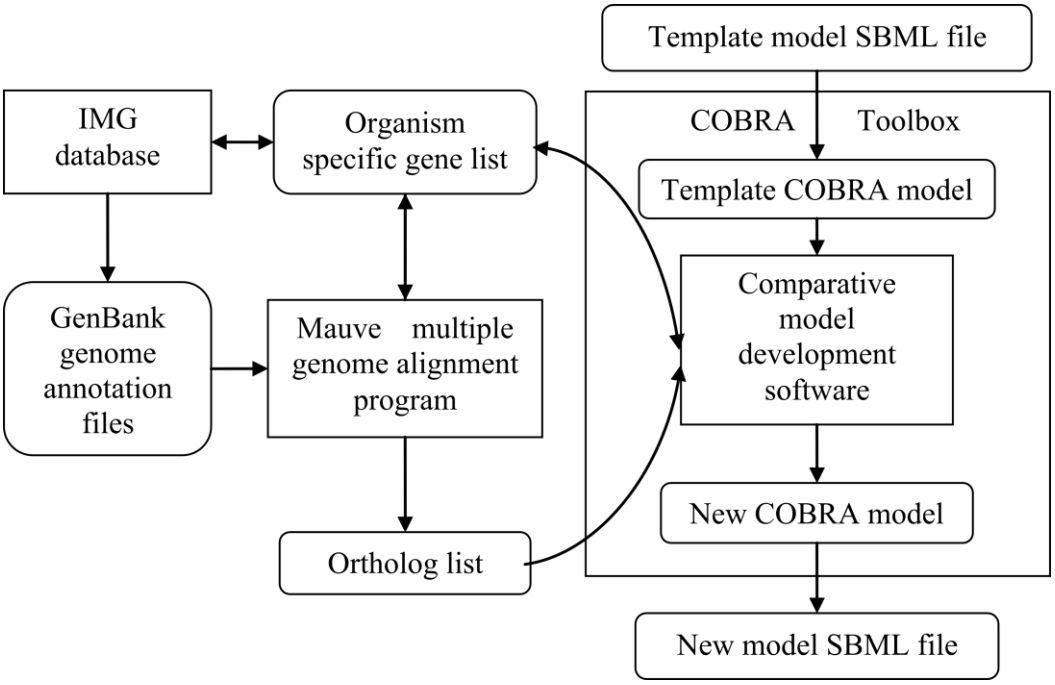


Figure 4.1 A flow diagram representing the comparative model reconstruction method. Data sources include GenBank files for *T. maritima* and *T. neapolitana* and the *T. maritima* (template) model in SBML format. The comparative model development software (Matlab code built upon COBRA and SBML toolbox code) generates a new model SBML file for *T. neapolitana* and two lists of genes, which are specific to each organism. The Mauve genome alignment and IMG database are used for manual curation of organism specific genes.

4.2.3 Flux Balance Analysis for Model Verification

To verify the functionality of the new model and identify differences between the two models, flux balance analysis (FBA) and dynamic flux balance analysis (dFBA) were performed. FBA is the use of linear optimization to calculate a steady-state flux distribution of the model network that gives a maximum (or minimum) solution to a particular objective function (Edwards et al. 2002). The most commonly used objective function for FBA is the maximization of biomass formation. This objective function was used for FBA with the *T. neapolitana* model on 60 potential carbon sources. The results were compared with the results Zhang et al. obtained with the *T. maritima* model (2009).

In dFBA, batch growth is simulated by solving the steady-state FBA optimization problem at discrete time steps for a given growth period (Varma and Palsson 1994; Becker et al. 2007). At each time step the maximum growth rate is determined and then used to calculate the quantity of biomass formed with the Monod equation. The new extracellular substrate and product concentrations are calculated using the biomass concentration and the reaction flux values. The calculated substrate concentrations determine the uptake flux for the model in the next time step. The COBRA toolbox dynamic flux balance analysis assumes exponential growth and that substrate concentration is the only limit to growth, however product inhibition of growth can be caused by elevated H_2 partial pressure (Schröder et al. 1994; van Niel et al. 2003). Therefore we tested the inclusion of a product inhibition term (Han and Levenspiel 1988), in the Monod equation used to estimate biomass formation. The specific growth rate, μ, with product inhibition is calculated from the following equation:

$$\mu = \mu_{\max}\left(1 - \frac{C_{H_2}}{C_{H_2,crit}}\right)^n \quad (4.1)$$

where

μ = specific growth rate (h^{-1}),

μ_{max} = maximum specific growth rate (h^{-1}),

C_{H_2} = inhibitor concentration ($mmol \cdot L^{-1}$),

$C_{H_2,crit}$ = critical inhibitor concentration at which growth ceases ($mmol \cdot L^{-1}$),

n = degree of noncompetitive inhibition, assumed to be 1.

An equation of this form was previously applied to evaluate product inhibition effects on H_2 production in *Caldicellulosiruptor saccharolyticus* (van Niel et al. 2003). The critical H_2 concentration of 27.5 mmol/L (per 50 mL of liquid culture volume) was estimated using the ideal gas law with a limiting H_2 partial pressure (P_{H_2}) of 0.3 atm at 25 °C and 110 mL headspace volume. Dynamic flux balance analysis with and without H_2 product inhibition was used to compare the newly developed *T. neapolitana* model with data from previous *T. neapolitana* batch growth experiments (Munro et al. 2009) and *T. maritima* data from Schröder et al. (1994).

4.2.4 Analysis of Species-Specific Genes

During the comparative model reconstruction the Mauve ortholog alignment file was processed to extract the locus tags that were identified by the alignment procedure as specific to each species. The gene lists were compared to the genes in the *T. maritima* and *T. neapolitana* models to eliminate those that had already been included in the models. Manual curation was then conducted with each gene list to identify any other instances of synteny. First, the genes were searched in the Mauve alignment program to identify any nucleotide sequence similarities. Product function assignments were obtained by a KEGG sequence similarity database search and by

parsing the GenBank files downloaded from IMG. The resulting lists were then searched in the IMG database for homology to other genome sequences. The *T. neapolitana* genes were further assessed to determine if any could be added to the model to reinstate the gene protein reaction associations (GPRs) that had been lost during the initial comparative model formation.

4.2.5 Model-Driven Hypothesis Generation

The *T. neapolitana* model was used to develop and test hypotheses related to cysteine and H_2 metabolism in this species. The model was first tested to determine if the addition of a cysteine uptake reaction would enable growth without elemental sulfur. It was confirmed that the model required elemental sulfur for growth, because no reactions were present in the network to convert cysteine to methionine and both are required for biomass formation. To identify possible reaction additions a subset network containing reactions related to elemental sulfur, cysteine, and methionine was isolated from the larger network. This network was compared with the *T. maritima*, *T. neapolitana,* and reference pathways (containing all reactions) in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2010). Possible reactions were identified that could enable growth on cysteine, without requiring elemental sulfur, and were added to the model. FBA simulations were conducted with the modified model and compared with results from the *T. neapolitana* model that could only utilize elemental sulfur as well as the original *T. maritima* model.

Recent experimental evidence and hypotheses from the literature related to hydrogenase activity in *Thermotoga* species were evaluated using FBA of the *T. neapolitana* model. Trimeric and monomeric hydrogenases were included with both solitary and synergistic electron carrier activities. Membrane protein activities

including ion gradient driven oxidoreductase and V-type ATPase activities were also tested in the model.

4.3 Results

4.3.1 *Thermotoga neapolitana* Central Carbon Metabolism Model

Using locus tag matches from the Mauve genome alignment program and manual curation a model for *T. neapolitana* was developed from a template *T. maritima* model. A list of genes that were manually curated is provided in Table 4.1. Manual curation of the model was necessary because of three different cases where there was not a clear one-to-one relationship between *T. neapolitana* and *T. maritima* genes. In the first case, two different gene annotations were provided for the same nucleotide sequence in the *T. neapolitana* genome annotation. When this occurred the two genes were treated as one gene in the model and the term 'DUAL' was used to indicate the dual annotation. Either annotation could match to the *T. maritima* sequence. The second case was of two ORF annotations covering a sequence in one genome that is covered by one ORF annotation in the other genome. Lastly there were several cases where the Mauve ortholog report indicated that no *T. neapolitana* ortholog was found for a gene present in the *T. maritima* model, but visual inspection of the annotated nucleotide sequences indicated that there was similarity between the nucleotide sequences of the two species.

The resulting *T. neapolitana* model has a high degree of similarity to the previously published *T. maritima* model. Only twenty five of the 479 genes in the *T. maritima* model did not have orthologs in the *T. neapolitana* genome sequence.

Table 4.1 Corresponding *T. maritima* and *T. neapolitana* gene annotations from manual curation using Mauve and the IMG database. Gene annotations separated by an underscore combine to produce the corresponding annotation. Annotations separated by DUAL are duplicate annotations of the same nucleotide sequence.

Corresponding Gene Annotations

T. maritima	***T. neapolitana***
TM0066	CTN_0627_CTN_0628
TM0111	CTN_0579DUALCTN_0580
TM0138	CTN_0552DUALCTN_0551
TM0155	CTN_0533
TM0292	CTN_0392DUALCTN_0391
TM0296	CTN_0387
TM0334	CTN_0334
TM0335	CTN_0334
TM0347	CTN_0323DUALCTN_0324
TM0356	CTN_0284_CTN_0285
TM0421	CTN_0249
TM0539	CTN_0129_CTN_0131
TM0689	CTN_1896_CTN_1897
TM0721	CTN_1863
TM0809	CTN_1768_CTN_1769
TM0891	CTN_1685
TM1067	CTN_1502_CTN_1503
TM1193	CTN_1381_CTN_1382
TM1418	CTN_1074_CTN_1075
TM1438	CTN_1055_CTN_1056
TM1596	CTN_0862_CTN_0863
TM1613	CTN_0845
TM1698	CTN_0885
TM1728	CTN_0925DUALCTN_0926
TM1836	CTN_0763
TM1837	CTN_0764
TM1875	CTN_0810DUALCTN_0811

These genes and twenty three dependent reactions were excluded from the *T. neapolitana* model. The reactions and the associated genes are shown in Table 4.2.

Several groups of *T. maritima* genes that were excluded from the *T. neapolitana* model were related to oligopeptide ABC transporters. These include: TM0300 – TM0304, TM1063 – TM1066, and TM1746 – TM1750. TM1226 represents an oligopeptide transporter periplasmic binding protein that has specificity for cellulose-derived oligomers (n = 4, 6), mannotriose, and mannotetraose. Three sequence annotations were correlated with glycerol-3-phosphate transport, TM1120 – TM1122, however neither the presence nor absence of these genes allows either model to grow on glycerol-3-phosphate.

Several genes related to hydrolytic activities were also excluded from the *T. neapolitana* model. TM1751 and TM1752 are associated with galactomannan and glucomannan hydrolysis. TM0305 is associated with glucomannan hydrolysis. TM0434 has been assigned alpha-glucosidase and alpha-glucoronidase activities and a *T. neapolitana* ortholog for this gene could not be found. However TM0055 can also function in this role and the *T. neapolitana* ortholog for TM0055 was included in the model; therefore the corresponding reaction was not removed.

Other genes excluded from the model included a pair of oxidoreductases associated with flavin adenine dinucleotide (FAD) (TM0427 and TM0428) and TM1429, which is associated with glycerol transport. The absence of the TM1429 sequence renders the *T. neapolitana* model incapable of growth on glycerol. The predicted inability of *T. neapolitana* to grow on glycerol is interesting since the highest H_2 yield for the *T. maritima* model was predicted with glycerol as a carbon source (Zhang et al. 2009). We attempted to grow *T. neapolitana* with glycerol as a carbon source and found that glycerol did not support growth. This result is in

Table 4.2 *T. maritima* model reactions and genes (bold-face) (Zhang et al. 2009) excluded from *T. neapolitana* model, [c] = cytoplasmic, [e] = extracellular.

Reactions Excluded from *T. neapolitana* model	**Gene Associations**
Glucomannan and Galactomannan (n = 4 and n = 6) transport via ABC transporter: $ATP[c] + H_2O[c] + Glucomannan(n)[e] \rightarrow ADP[c] + Glucomannan(6)[c] + H^+[c] + P_i[c]$ $ATP[c] + H_2O[c] + Galactomannan(n)[e] \rightarrow ADP[c] + Galactomannan(n)[c] + H^+[c] + P_i[c]$	(**TM1746** and **TM1747** and **TM1748** and **TM1749** and **TM1750**)
β-1,3/1,4-Glucan (n = 4 and n = 6) transport via ATP transporter: $ATP[c] + H_2O[c] + \beta\text{-}1,3/1,4\text{-}glucan(n)[e] \rightarrow ADP[c] + \beta\text{-}1,3/1,4\text{-}glucan(n)[c] + H^+[c] + P_i[c]$	(**TM0300** and **TM0301** and **TM0302** and **TM0303** and **TM0304**)
Cellulose and Mannose oligomer transport via ABC transporter: $ATP[c] + H_2O[c] + mannotriose [e] \rightarrow ADP[c] + H^+[c] + mannotriose[c] + P_i[c]$ $ATP[c] + H_2O[c] + mannotetraose[e] \rightarrow ADP[c] + H^+[c] + mannotetraose[c] + P_i[c]$ $ATP[c] + H_2O[c] + Oligo\text{-}cellulose(n)[e] \rightarrow ADP[c] + Oligo\text{-}cellulose(n)[c] + H^+[c] + P_i[c]$	(**TM1226** and TM1222 and TM1221 and TM1220 and TM1219)
L-Rhamnose transport (ATP-dependent): $ATP[c] + H_2O[c] + L\text{-}rhamnose[e] \rightarrow ADP[c] + H^+[c] + P_i[c] + L\text{-}rhamnose[c]$	(TM1067 and **TM1066** and **TM1065** and **TM1064** and **TM1063**)
Glycerol 3-phosphate transport via ABC transporter: $ATP[c] + H_2O[c] + Glycerol\text{-}3\text{-}P[e] \rightarrow ADP[c] + Glycerol\text{-}3\text{-}P[c] + H^+[c] + P_i[c]$	(**TM1121** and **TM1120** and **TM1122**)
Glycerol transport via uniport (facilitated diffusion): $Glycerol[e] \rightarrow Glycerol[c]$	**TM1429**
Hydrolysis of galactomannan (n = 4 and n = 6): $Galactomannan(4)[c] + 3\ H_2O[c] \rightarrow Galactose[c] + 3\ Mannose[c]$ $Galactomannan(6)[c] + 5\ H_2O[c] \rightarrow 2\ Galactose[c] + 4\ Mannose[c]$	(**TM1751** and **TM1752** and TM1192 and TM1624)
Hydrolysis of glucomannan: $Glucomannan(600)[e] + 149\ H_2O[e] \rightarrow 150\ Glucomannan(4)[e]$ $Glucomannan(600)[e] + 99\ H_2O[e] \rightarrow 100\ Glucomannan(6)[e]$ $Glucomannan(600)[e] + 124\ H_2O[e] \rightarrow 75\ Glucomannan(4)[e] + 50\ Glucomannan(6)[e]$ $Glucomannan(6)[c] + 5\ H_2O[c] \rightarrow 2\ Glucose[c] + 4\ Mannose[c]$ $Glucomannan(4)[c] + 3\ H_2O[c] \rightarrow Glucose[c] + 3\ Mannose[c]$	(**TM0305** and TM1227) (**TM1751** and **TM1752** and TM1624)
Oxidoreductase activity: $FAD[c] + 3\ H^+[c] + NADPH[c] \leftrightarrow FADH_2[c] + NADP[c] + 2\ H^+[e]$ $FAD[c] + Succinate[c] \leftrightarrow FADH_2[c] + Fumarate[c]$ $FADH_2[c] + Quinone[c] \leftrightarrow FAD[c] + QuinoneH_2[c]$	(**TM0428** and **TM0427**)

agreement with previous experimental studies (Belkin et al. 1986; Jannasch et al. 1988).

4.3.2 Model Verification: FBA and dFBA

To verify the use of the *T. maritima* model to construct the *T. neapolitana* model both FBA and dFBA were performed. FBA was conducted using 60 carbon sources with both the *T. maritima* and *T. neapolitana* models. The results for maximum biomass were compared to the results presented by Zhang et al. for *T. maritima* in their supplementary files (2009). Fluxes were identical for 44 of the 60 substrates. For the remaining model carbon sources no growth was predicted with the *T. neapolitana* model, but growth was predicted in the *T. maritima* model. Twelve of the sixteen model substrates represented polymers: cellulose (n = 4, 6, and 300 repeat units), galactomannan (n = 4, 6, and 600), glucomannan (n = 4, 6, and 600), β-1,3/1,4-glucan (n = 4, 6, and 1500). The remaining four substrates that did not support growth were mannotriose, mannotetraose, L-rhamnose, and glycerol. Cellotetraose (a cellulose polymer with n = 4 repeat units of glucose), glycerol, and L-rhamnose were each experimentally tested as carbon sources for *T. neapolitana*, and only cellotetraose supported cell growth. Glycerol and L-rhamnose were not metabolized by *T. neapolitana*. Genes specific to *T. neapolitana* that are not found in *T. maritima* were analyzed to identify any ORFs that might be included in the model to allow growth on cellotetraose and the other substrates that have not yet been experimentally tested.

The *T. neapolitana* model dFBA predictions were compared to experimental data from *T. neapolitana* (Munro et al. 2009) and *T. maritima* (Schröder et al. 1994). The results of the model prediction and experimental results are presented in

Figure 4.2. Data points for *T. neapolitana* batch growth prior to 12 hours were not available; however data for *T. maritima* under similar growth conditions during the first 12 hours of growth indicates similar cell growth (Figure 4.2A) and metabolite profiles (data not shown). The increased growth rate for *T. maritima* is due to an increased initial biomass concentration in those experiments (5 mg CDW/L) compared to the initial biomass concentration in the *T. neapolitana* experiments (1.2 mg CDW/L). The inclusion of H_2 inhibition in the dynamic flux balance analysis improved the model prediction of when growth would cease and the maximum biomass concentration (Figure 4.2A). In Figure 4.2B the dFBA product inhibition model predictions for H_2, acetate and glucose match well with the experimentally observed *T. neapolitana* metabolite profiles.

4.3.3 Analysis of Species-Specific Gene Lists

The full species-specific gene lists for *T. maritima* and *T. neapolitana* are in Appendix 2. Ortholog neighborhood analysis using IMG revealed that there were groups of carbohydrate transporter genes with no assigned function found in the *T. neapolitana* specific gene group. Several of these CUT1 family carbohydrate ABC transporter clusters are listed in Table 4.3. It is possible that these genes are associated with the transport reactions in *T. maritima* that were eliminated from the *T. neapolitana* model (Table 4.2) although this needs experimental verification.

4.3.4 Cysteine Metabolism

The *T. maritima* model requires an uptake reaction for elemental sulfur to simulate growth on a defined glucose minimal medium. This accurately reflects

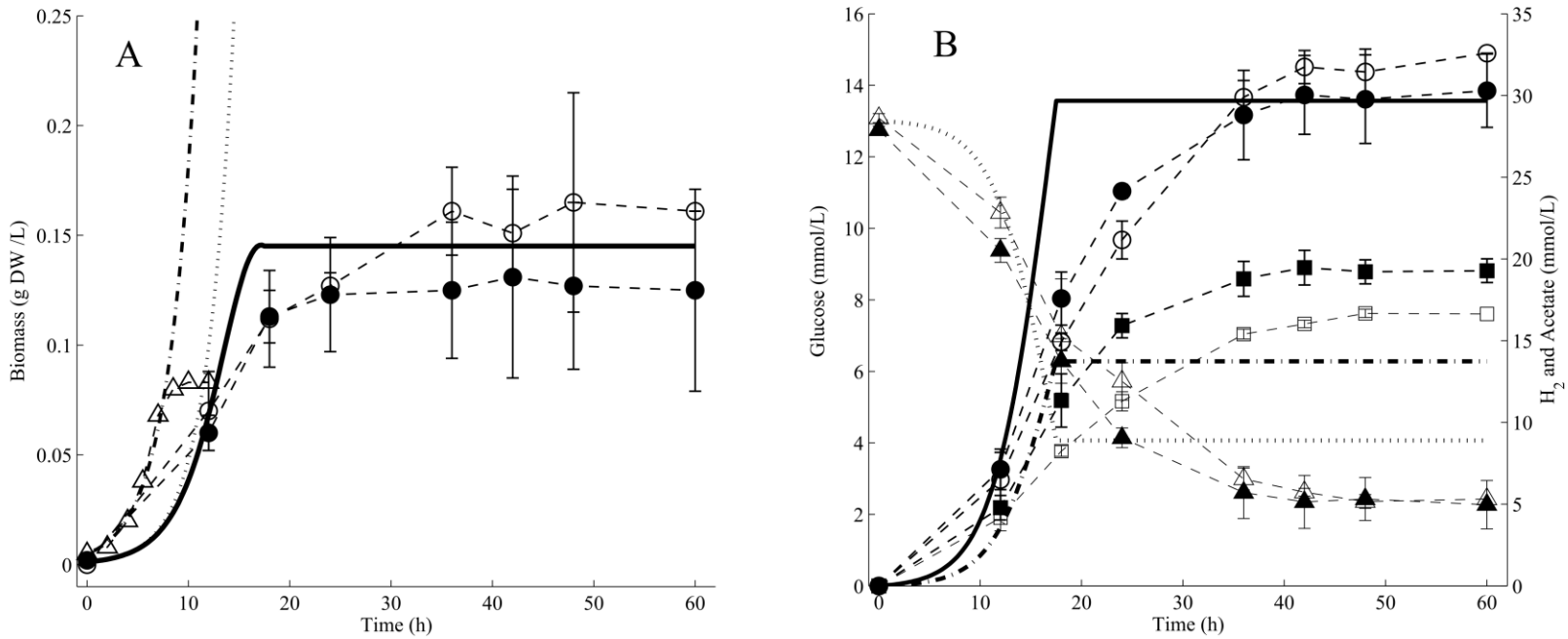


Figure 4.2 Dynamic flux balance analysis with the *T. neapolitana* model compared to revised experimental data from a previous study (Munro et al. 2009) with *T. neapolitana* and a growth study with *T. maritima* (Schröder et al. 1994) **A)** Biomass produced in 85 °C (●) and 77 °C (○) *T. neapolitana* batch cultures and *T. maritima* batch cultures (△), dFBA results are shown for exponential growth in *T. maritima* (- · - ·) and *T. neapolitana* (·····) and product inhibition growth in *T. neapolitana* (——) **B)** *T. neapolitana* results for H_2 at 85 °C (●) and 77 °C (○), acetate at 85 °C (■) and 77 °C (□) , and glucose at 85 °C (▲) and 77 °C (△) dFBA product inhibition results are shown for H_2 (——), acetate (- · - ·) and glucose (·····). Lines through data points are guides for the eye.

Table 4.3 *T. neapolitana* specific genes have been annotated as CUT1 family carbohydrate transporters which consist of extracellular solute-binding proteins and binding-protein-dependent transport systems. Ortholog neighborhood analysis was conducted in IMG to find similar annotations in different species.

CUT1 Family Transporters			
Extracellular solute-binding protein	Binding-protein-dependent transport systems		Ortholog Neighborhood Analysis
CTN_0660	CTN_0661	CTN_0662	Similar to *Kosmotoga olearia, Thermosipho,* and *Petrotoga* species
CTN_0780	CTN_0778	CTN_0779	Similar to *Thermotoga* sp. RQ2
CTN_1372	CTN_1373	CTN_1374	Similar to *Fervidobacterium nodosum, Thermotoga lettingae, Halothermothrix orenii*
CTN_1541	CTN_1542	CTN_1543	No orthologs identified

experimental results that indicate the inability of *T. maritima* to grow without either elemental sulfur or a complex medium component such as yeast extract (Schröder et al. 1994). *T. neapolitana* on the other hand is capable of growth on glucose minimal medium without elemental sulfur, because cysteine- serves as both a reducing agent and a sulfur source (Childers et al. 1992). Both cysteine and methionine are components of the biomass formation reaction; therefore KEGG was used to develop enzyme pathways that could allow *T. neapolitana* to form methionine from cysteine. The proposed model pathways for sulfur and cysteine metabolism are shown in Figure 4.3. A cysteine uptake reaction was included in the model and it was assumed to be ATP-dependent like the other amino acid transporters in the model. Cystathionine-γ-synthase (E.C. 2.5.1.48) activity was already present in the *T. neapolitana* model and either CTN_1303 or CTN_1694 can be associated with this protein. Cystathionine-β-lyase (E.C. 4.1.1.8) cleaves cystathionine to form homocysteine which enables the formation of methionine from cysteine without formation of H_2S. However, a distinct gene annotation encoding cystathionine-β-lyase could not be identified in the *T. neapolitana* genome. Therefore there is still no explanation for why *T. neapolitana* can and *T. maritima* cannot use cysteine as a sulfur source. The model was tested with a cystathionine-β-lyase reaction and unrestricted uptake bounds for both elemental sulfur and cysteine fluxes. The optimal FBA solution for maximum growth relied on cysteine as a sulfur source; no elemental sulfur was consumed when cysteine reactions were also made available. There was no significant change in the predicted maximum growth rate, $\mu = 0.36\ h^{-1}$, for the *T. neapolitana* model with both cysteine and elemental sulfur pathways and the model version that was dependent on elemental sulfur.

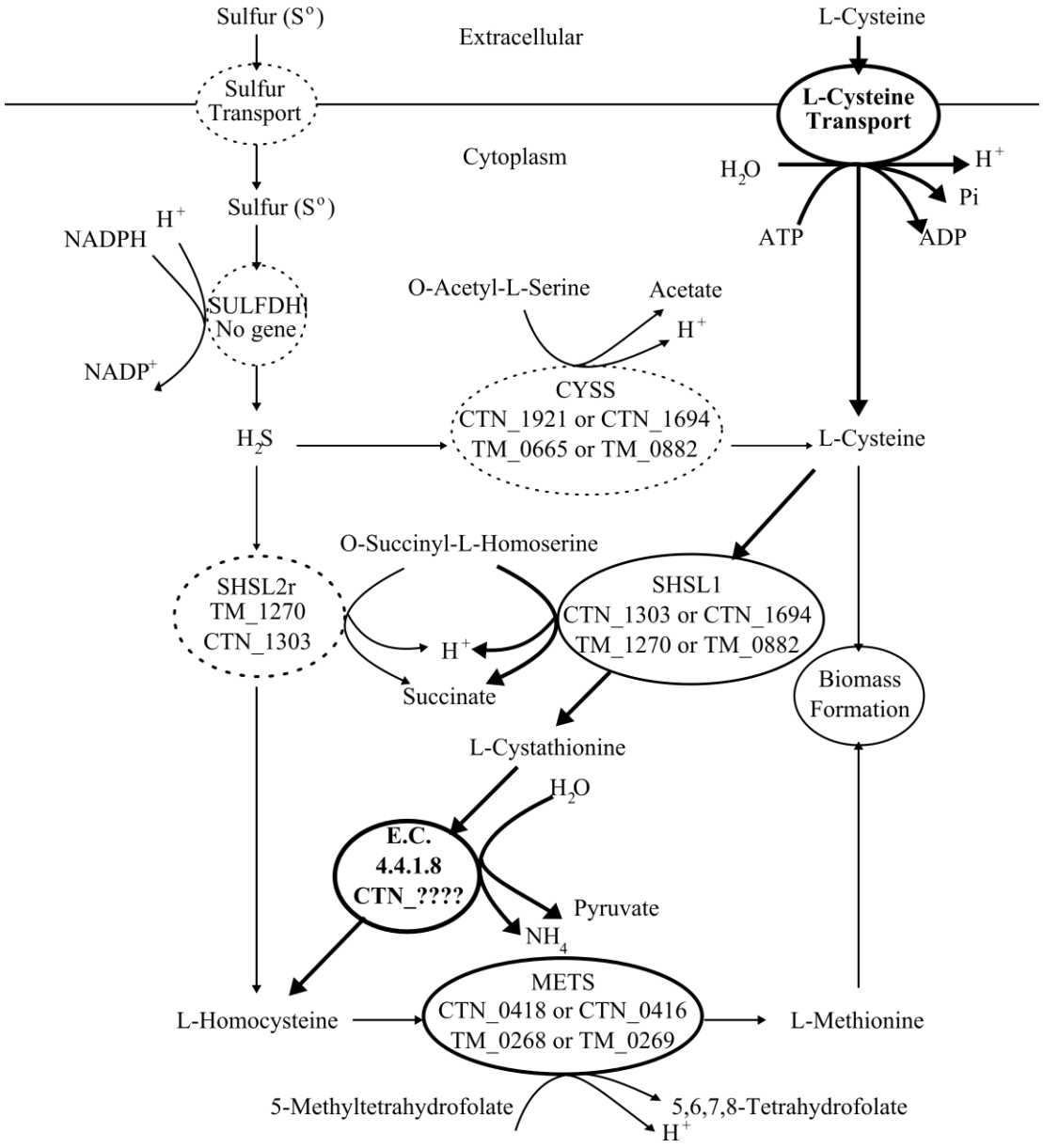


Figure 4.3 Subset network containing elemental sulfur, cysteine, and methionine. Dashed lines represent network reactions that occur when elemental sulfur is the sole sulfur source, these include: Sulfur transport, sulfide dehydrogenase (SULFDH), cysteine synthase (CYSS), and O-succinylhomoserine lyase (H_2S specificity) (SHSL2r). Bold lines represent L-cysteine transport and cystathionine-β-lyase (E.C. 4.4.1.8) activities which are hypothesized to occur when cysteine is the sulfur source. O-succinylhomoserine lyase (SHSL1) is in the network for both conditions, but is only active when L-cysteine is provided as a sulfur source. Biomass formation and methionine synthesis (METS) occur under both conditions.

4.3.5 H_2 Reaction Modifications

The *T. neapolitana* model was tested with FBA to determine the effects of modifying reactions relevant to H_2 production to reflect new experimental data and hypotheses from the literature. The FBA simulation results are shown in Table 4.4 and the original and updated networks related to H_2 production are shown in Figure 4.4. In the original *T. maritima* model two separate reactions are identified for H_2 formation using reduced ferredoxin (HYDFDi) and NADH (NAD_H2). It has been shown recently in the literature that the trimeric hydrogenase in *T. maritima* uses ferredoxin (Fd) and NADH synergistically and a second monomeric hydrogenase uses Fd alone (Schut and Adams 2009). Genes for both hydrogenases are present in *T. neapolitana*; therefore the *T. neapolitana* model was updated to include both HYDFDi and a bifurcating hydrogenase activity (BifHYD). FBA with the modified model revealed that a separate flux through the NADH hydrogenase (NAD_H2) was necessary to maintain the correct NAD^+/NADH ratio for biosynthesis. In biochemical analysis of hydrogenase activities in *T. maritima* no NADH specific hydrogenase has been found (Schröder et al. 1994), therefore an alternative mechanism must be responsible for NADH cycling. One potential mechanism is a putative membrane bound NADH:Fd oxidoreductase with ion translocating activity that has been annotated in both the *T. maritima* and *T. neapolitana* sequences. Preliminary experiments with the enzyme from *T. maritima* have shown high NADH oxidation activity although the complete reaction mechanism has not been determined (Schut and Adams 2009). Proton translocation could drive the oxidoreductase reaction (Figure 4.4B), but it could also be driven by Na^+ flux into the cytoplasm (Figure 4.4C) (Saeki and Kumagai 1998; Boiangiu et al. 2005; Backiel et al. 2008).

Table 4.4 FBA results for the base model (TNA) and model iterations with different reactions for hydrogenase, oxidoreductase (OR), and ATPase activities. The carbon source for all cases was glucose = 10 mmol/g DW/h.

	FBA Solution Reaction Fluxes (mmol/g DW/h)							
Reactions	**TNA**	**3HYD**	**3HYD OR-H**	**3HYD OR-Na**	**2HYD**	**2HYD OR-H**	**2HYD OR-Na**	**2HYD OR-Na ATP**
Biomass Formation:	0.36	0.36	0.42	0.36	0	0.36	0	0.37
Acetate Production:	15.38	15.38	14.54	15.38	20	15.43	20	15.29
CO_2 Production:	16.46	16.46	15.81	16.46	20	16.49	20	16.38
H_2 Production:	33.22	33.22	31.98	33.22	40	33.28	40	33.08
Fd-only Hydrogenase: fdxr-4-2[c] + 2h[c] -> fdxo-4-2[c] + h2[c]	16.03	16.03	0	16.03	0	0	0	1.31
NADH Hydrogenase: h[c] + nadh[c] <=> h2[c] + nad[c]	17.19	17.19	31.98	17.19	--	--	--	--
Bifurcating Hydrogenase: fdxr-4-2[c] + nadh[c] + 3h[c] -> fdxo-4-2[c] + 2h2[c] + nad[c]	--	0	0	0	20	16.64	20	15.88
H^+ NADH Fd OR: fdxo-4-2[c] + nadh[c] + h[c] + 2h[e] <=> fdxr-4-2[c] + nad[c] + 2h[c]	--	--	-15.31	--	--	0.57	--	--
Na^+ NADH Fd OR: fdxo-4-2[c] + nadh[c] + h[c] + na1[e] <=> fdxr-4-2[c] + nad[c] + na1[c]	--	--	--	0	--	--	0	1.25
V-Type ATPase: atp[c] + na1[c] <=> adp[c] + pi[c] + na1[e]	--	--	--	--	--	--	--	1.25

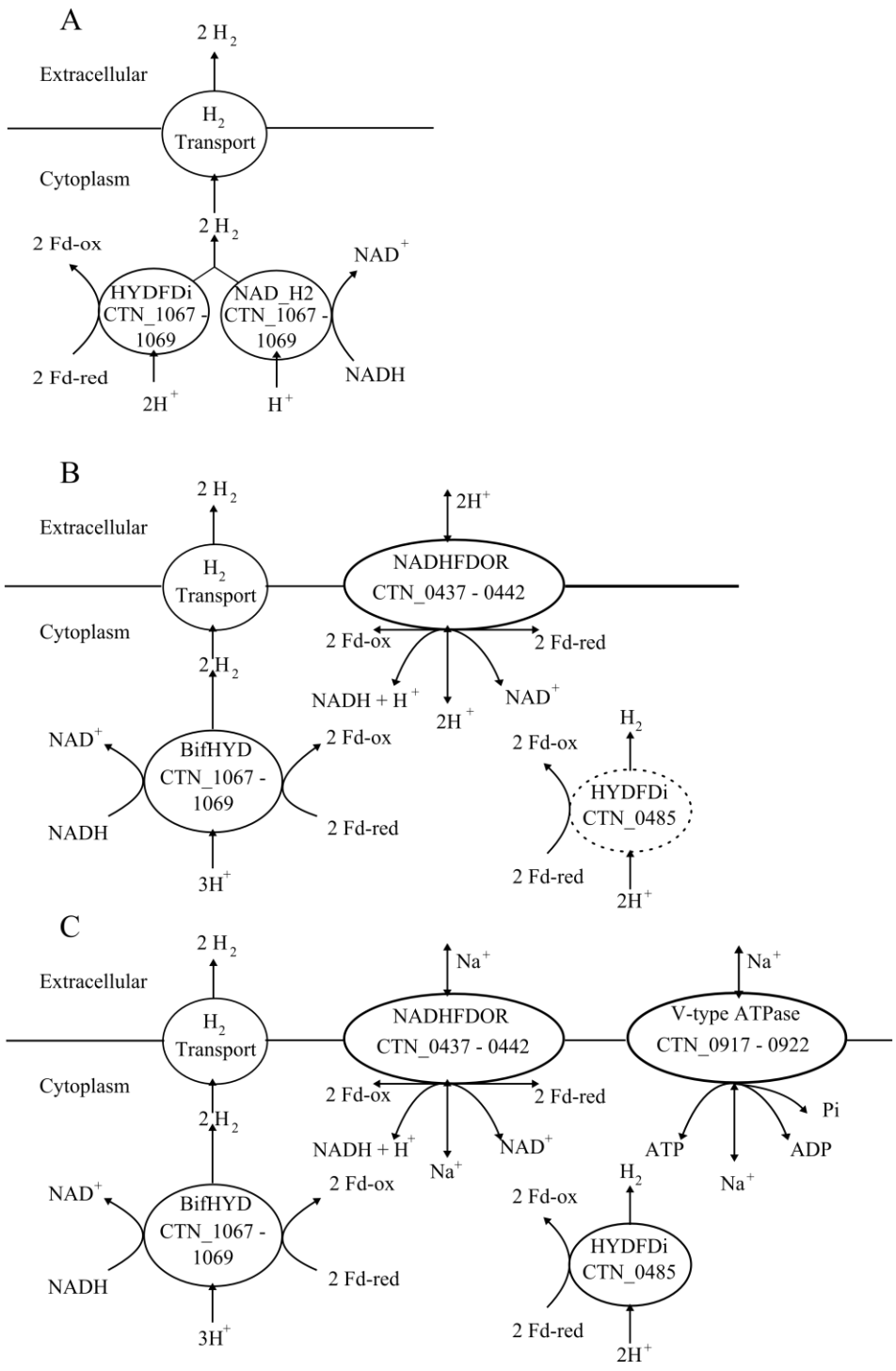


Figure 4.4 Subset networks for H_2 metabolism. **A)** Reactions in original model **B)** Reactions in the model with bifurcating hydrogenase activity (BifHYD) and a proton driven NADH:Ferredoxin oxidoreductase (NADHFDOR), the Fd-only hydrogenase (HYDFDi) is inactive in this network **C)** Reactions in the model with a sodium driven NADH:Ferredoxin oxidoreductase and V-type ATPase. The HYDFDi reaction is active in this network.

The only sodium transport reaction in the original *T. maritima* model is an irreversible sodium import reaction that is inactive since there are no other sodium related reactions in the model to balance it. Both *Thermotoga* species grow in high NaCl concentrations, and in addition to proton exchange these species may rely upon Na^+ driven transport activities. For example, Galperin et al. (1996) obtained results indicating that glucose transport in *T. neapolitana* was dependent on ATP and a Na^+ gradient.

Given that *T. neapolitana* could have a Na^+ gradient driven membrane bound NADH:Fd oxidoreductase, a reaction representing this activity was tested in the *T. neapolitana* model. The reaction could replace the NADH hydrogenase activity if the pre-existing sodium transport reaction in the model (NAti) was allowed to be reversible. However, the flux balance solution for this model version resulted in a significantly higher growth rate ($\mu = 0.48\ h^{-1}$) and also only produced H_2 through the ferredoxin specific hydrogenase, requiring high Na^+ ion flux and completely bypassing the bifurcating hydrogenase. Another alternative balancing reaction for sodium transport that could be included in the *T. neapolitana* model is a putative Na^+ driven V-type ATPase identified in the list of genes specific to *T. neapolitana* (not present in the *T. maritima* genome) (Appendix 2). Ortholog neighborhood analysis of these genes (CTN_0917 – CTN_0922) revealed similarity to both *Clostridium tetani* and *Halothermothrix orenii* V-type ATPase gene clusters. The V-type ATPase in *Clostridium tetani* has been identified as an Na^+ pump and drives many other Na^+ related transport reactions in this species (Bruggemann et al. 2003). Considering that many genes in the *Thermotogales* are closely related to those from *Clostridia* (Zhaxybayeva et al. 2009) the presence of this common gene cluster is not surprising. When this Na^+ driven V-type ATPase was added to the model and the Na^+ transport was converted to its original irreversible form (2HYDNaATP in Table 4.4), the model

produced H_2 through the bifurcating hydrogenase with a small amount through the ferredoxin hydrogenase (a 16:1 ratio). The Na^+ oxidoreductase and V-type ATPase reactions were the only active Na^+ reactions and flux values were significantly smaller than in the previous test.

4.4 Conclusions

We have demonstrated a method to generate a constraint-based model for one species from an existing, highly detailed model of another species and a genome alignment between the two annotated genomes. This strategy is advantageous, because it extends the utility of the original model reconstruction (Zhang et al. 2009) created by other researchers who expended considerable effort (Thiele and Palsson 2010). The multiple genome alignment program, Mauve, was integral to this procedure, because of its capacity for processing sequences with large-scale inversions, which is important for analysis of species with high similarity and large-scale rearrangements like *T. maritima* and *T. neapolitana*.

FBA growth simulations on various carbon sources showed that for most cases the *T. maritima* and *T. neapolitana* models have identical growth capabilities. However, several carbon sources could not support growth for the *T. neapolitana* model due to the absence of genes corresponding to the required reactions. Ortholog neighborhood analysis of the *T. neapolitana* specific gene list did not reveal any genes that could definitively reassign these reactions to the model. Although several CUT1 family transporters were present in the *T. neapolitana* specific gene list the specificity for these transporters is unknown. Experimental validation will be needed to identify the substrate specificity of these transport proteins and to verify the *T. neapolitana* model substrate utilization predictions.

The *T. neapolitana* model was used to generate hypotheses related to metabolism in *T. neapolitana* and *T. maritima*. Possible reactions that occur in *T. neapolitana,* but not in *T. maritima,* for uptake of cysteine and conversion to methionine were tested with the *T. neapolitana* model. We tested the inclusion of a cystathionine-β-lyase activity in the *T. neapolitana* model, but could not conclude from our analysis how *T. neapolitana* is capable of using cysteine as a sole sulfur source.

Reactions associated with H_2 production that are likely present for both species were also tested in the *T. neapolitana* model. A newly identified bifurcating hydrogenase activity (Schut and Adams 2009) was added to the model and the NADH hydrogenase activity was removed. The incorporation of the bifurcating hydrogenase required the addition of NADH oxidizing oxidoreductase activity. Since a putative ion translocating membrane bound oxidoreductase is in the genome annotations of both species it was tested in the *T. neapolitana* model. Both proton and sodium ion translocating oxidoreductases were tested in the model. The H^+ translocating oxidoreductase supported growth without including additional reactions. The Na^+ translocating oxidoreductase required a balancing sodium efflux reaction. A Na^+ dependent V-type ATPase found in *T. neapolitana* (but not in *T. maritima*) supported growth and fluxes through the bifurcating and Fd-only hydrogenase reactions. These results indicate that *T. neapolitana* could use either proton or sodium-driven reactions to support growth and hydrogen production, but to verify such activity in either species as well as the other hypotheses generated from our analysis, future experimental work will be necessary.

Chapter 5

Experimental Testing of Hypotheses Derived from the *Thermotoga neapolitana* Constraint-Based Model

ABSTRACT

Carbon source utilization predictions derived from a constraint-based model of *T. neapolitana* central carbon metabolism were experimentally tested. The model predicted that *T. neapolitana* could not grow on glycerol, L-rhamnose, or cellotetraose. Batch growth experiments show that glycerol and L-rhamnose are not metabolized by *T. neapolitana*; growth observed in these cultures is due to metabolism of yeast extract. Cellotetraose did support growth. Rates of cellotetraose consumption were comparable to cellobiose consumption however cell yields per glucose equivalent were higher on cellotetraose than on cellobiose or glucose. Proteomic analysis showed that several proteins were differentially expressed when grown on cellotetraose and glucose.

5.1 Introduction

In a previous study (see Chapter 4), a constraint-based model representing the central metabolic network of *Thermotoga neapolitana* was created with a novel comparative model reconstruction method. In this method an existing model of *T. maritima* central metabolism (Zhang et al. 2009) was translated into a model for *T. neapolitana* using the Mauve genome alignment program (Darling et al. 2004) and the published genome annotation for *T. maritima* (Nelson et al. 1999) and

T. neapolitana (Lim et al. 2009). The method relied on internally developed as well as open source computational methods in Matlab (Keating et al. 2006; Becker et al. 2007; Makhorin 2008). The model was used to develop hypotheses related to carbon source degradation and sulfur metabolism. Experimental evaluation of model hypotheses is crucial to improve the accuracy of the model. In this chapter experimental testing of model-derived hypotheses related to carbon source metabolism is described.

Flux balance analysis of the *T. neapolitana* model predicted that this species could not grow on several carbon sources including cellotetraose, glycerol, and L-rhamnose. There is conflicting evidence in the literature related to glycerol metabolism in this species (Belkin et al. 1986; Jannasch et al. 1988; Van Ooteghem et al. 2004). We show here that glycerol and L-rhamnose are not metabolized by *T. neapolitana* cells; growth that was observed on these substrates can be attributed to metabolism of yeast extract. The model predicted that *T. neapolitana* could not grow on cellotetraose. However, *in vitro* enzyme activity has been shown on cellotetraose with cellobiose phosphorylase and β-glucosidase enzymes cloned from *T. neapolitana* cells grown on cellobiose (Yernool et al. 2000). The experimental data presented in this study is the first demonstration that cellotetraose can support *T. neapolitana* growth *in vivo*. Proteomic analysis of cells grown on cellotetraose and glucose was used to identify proteins associated with growth on cellotetraose.

5.2 Methods

5.2.1 Microorganism and Growth Medium

T. neapolitana cultures obtained from the DSMZ culture collection were grown as described previously under anaerobic conditions in 160 mL serum bottles

filled with 50 mL of *Thermotoga* basal (TB) medium (Childers et al. 1992). The growth medium was titrated with NaOH to obtain an initial pH of 7.5 for all experiments. Culture medium was bubbled with nitrogen (N_2) for 15 min in the serum bottles to create anaerobic atmospheres. The bottles were then sealed with butyl rubber stoppers and aluminum crimp caps and autoclaved. The TB medium was either supplemented with glucose (TBG), glucose and yeast extract (TBYG), cellobiose and yeast extract (TBYC2), cellotetraose and yeast extract (TBYC4), glycerol and yeast extract (TBYGly) or L-rhamnose and yeast extract (TBYR). Yeast extract was added from a separate anoxic stock solution (10% w/v) for a final concentration of 0.5 $g \cdot L^{-1}$. To achieve the carbon source concentrations listed in Table 5.1, 0.5 mL of 100x concentrated carbon substrate stock solutions were added to 50 mL medium using sterile syringe filtration techniques. Stock solutions were injected into the sterilized culture bottles through a 0.2 µm sterile filter attached to a sterile syringe flushed with N_2. Cultures were inoculated with a 1% v/v inoculum from a TBYG culture and unless otherwise specified, grown in an orbital shaking water bath (Gyrotory Shaker, New Brunswick Scientific, Edison, NJ) set at 75 rpm and 77 °C.

Table 5.1 Growth medium formulations for 30 h batch growth experiments.

Growth Medium	Carbon Source	Initial Concentration (mM)	Glucose Equivalents (mM)	Yeast Extract
TBYG	Glucose	4.8	4.8	+
TBG	Glucose	4.6	4.6	-
TBYC2	Cellobiose	1.5	2.9	+
TBYC4	Cellotetraose	0.3	1.3	+
TBYGly	Glycerol	10.2	5.1	+

5.2.2 Batch Growth Experiments

Batch growth experiments were conducted for a 30 h period to quantify growth of *T. neapolitana* on different carbon and sulfur sources. The initial conditions for these experiments are listed in Table 5.1. Duplicate cultures for each condition were prepared and sampled at the following intervals during the 30 h growth period: 0, 4, 8, 12, 18, 24, and 30 h. At each time point, 0.5 mL of headspace volume was removed for H_2 analysis and 1.7 mL of liquid volume was removed for turbidity measurements, Bradford total protein analysis, and HPLC analysis of carbon source consumption and organic acid production.

An additional experiment was conducted to compare growth on TBYG medium and L-rhamnose medium (TBYR). In this study cultures were grown in an oven (Barnstead Thermolyne, Dubuque, IA) at 77 °C. Samples were removed for analysis at 0 and 24 h of growth. The same metabolite and growth measurements were made as for the 30 h culture experiments. The initial glucose concentration and initial L-rhamnose concentrations were 6.0 mM glucose and 5.5 mM L-rhamnose.

5.2.3 Fermentation Metabolite and Growth Analysis

For H_2 analysis 0.5 mL headspace samples were removed from each culture using a gastight syringe (Hamilton or Henke Sass Wolf GmbH). Hydrogen production was measured with a gas chromatograph (model 310, SRI Instruments, Torrance, CA) equipped with a thermal conductivity detector and a 0.9 m Molesieve 5A column. The carrier gas was N_2 and the column oven and detector temperatures were 27 and 100 °C respectively. Calibration curves were generated using H_2 standards to measure the

partial pressure of H_2 in the bottle headspace. The H_2 partial pressure measurements were used to estimate the mmol of H_2 produced per liter liquid culture volume.

Liquid samples were analyzed for organic acids, glucose, cellotetraose, cellobiose, L-rhamnose, and glycerol with a Shimadzu HPLC system (Shimadzu, Columbia, MD) equipped with a refractive index detector. For analysis of organic acids a Bio-Rad Aminex HPX-H column was used. The system was operated at 25 °C with a mobile phase of 5 mM H_2SO_4 at a flow rate of 0.6 $mL \cdot min^{-1}$. Glucose, cellotetraose, cellobiose, L-rhamnose, and glycerol were quantified with a BioRad Aminex HPX-P column. The system was operated at 85 °C with a mobile phase of H_2O and a flow rate of 0.55 $mL \cdot min^{-1}$. Standard curves were prepared and used for quantification of all measured metabolites.

Cell growth was quantified by absorbance measurements at 670 nm and the modified Bradford protein assay (Stoscheck 1990).

5.2.4 Sample Preparation for Proteomic Analysis

Duplicate cultures were grown with TBG, TBYG, and TBYC4 medium. After 18 h of growth, incubation was terminated and cultures were placed on ice. Samples were withdrawn for H_2 and organic acid analysis and 25 mL of culture volume was centrifuged at 8,000 g and 4 °C for 30 min. Culture supernatant and cell samples were then separated and frozen at -80 °C for proteomic analysis by researchers in Kelvin Lee's laboratory at the University of Delaware.

5.2.5 Proteomic Analysis

All proteomic analysis was conducted at the University of Delaware by researchers in Kelvin Lee's laboratory. Two dimensional protein electrophoresis (2DE) of glucose and cellotetraose grown cells was conducted as described previously (Hatzimanikatis et al. 1999). Spots that were differentially expressed between the two experimental conditions were selected and analyzed by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS) as described by Finehout and Lee (2003). Mascot (v2.2. from MatrixScience) was used for the identity search of the mass spectrometry results in the NCBInr database. The following search parameters were used: bacterial taxonomy, trypsin enzyme, 1 missed cleavage allowed, 50 ppm protein mass tolerance, and 0.3 Da peptide mass tolerance.

5.3 Results

5.3.1 Batch Growth Results

Multiple carbon sources were tested to validate model predictions of carbon substrate utilization. Absorbance and total protein measurements to quantify growth in cultures with glucose, cellobiose, cellotetraose, and glycerol are shown in Figure 5.1 A and B. Substrate utilization results are shown in Figure 5.2. Substrate concentrations were normalized to mmol glucose equivalents per L culture by multiplying the molar concentration of a compound by the molecular weight of the compound divided by the molecular weight of glucose. The time profiles of the major fermentation products H_2 and acetate are shown in Figure 5.3A and B.

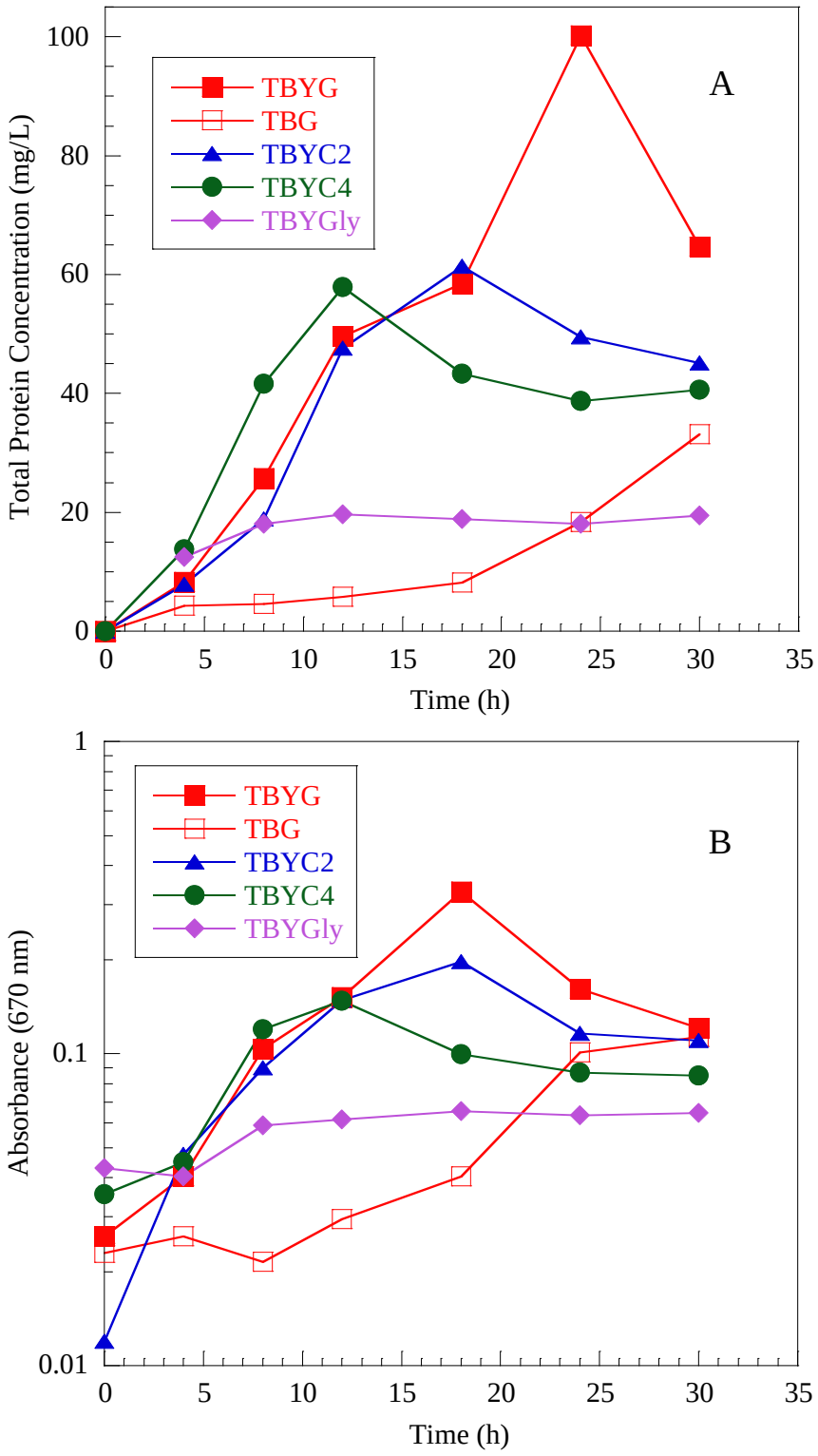


Figure 5.1 Protein concentration from the modified Bradford assay (A) and Absorbance (670 nm) (B) measurements as a function of time for different growth medium formulations. Data points represent averages of duplicate cultures.

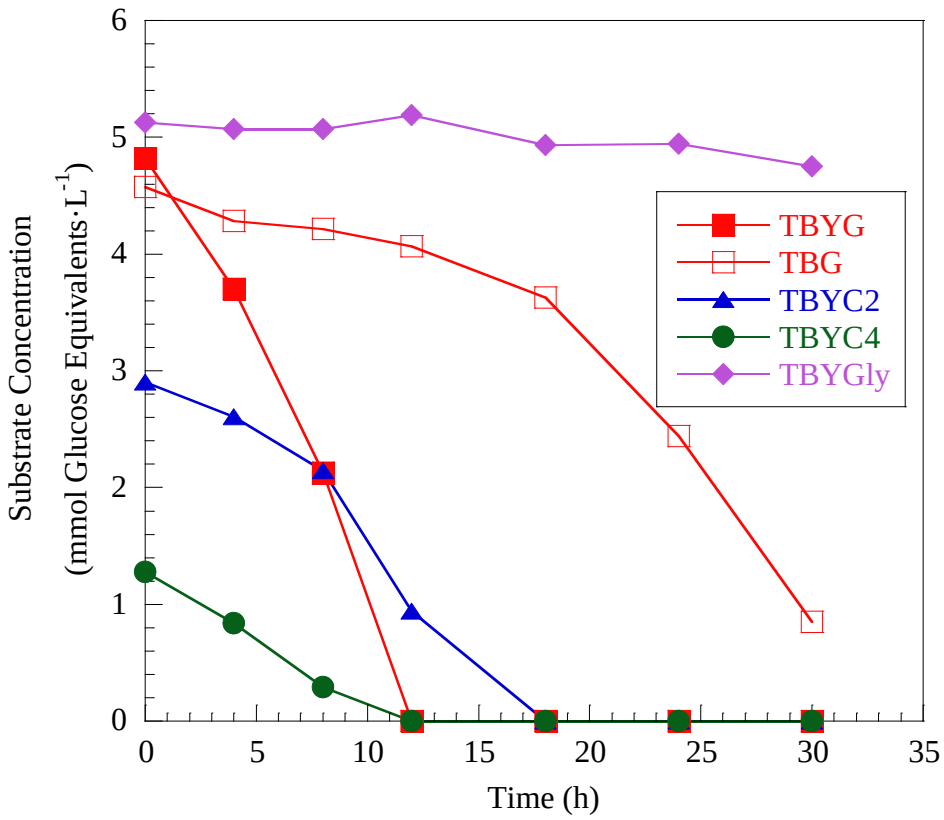


Figure 5.2 Substrate concentration (in mmol Glucose Equivalents per L culture) as a function of time for different growth medium formulations. Data points represent averages of duplicate cultures.

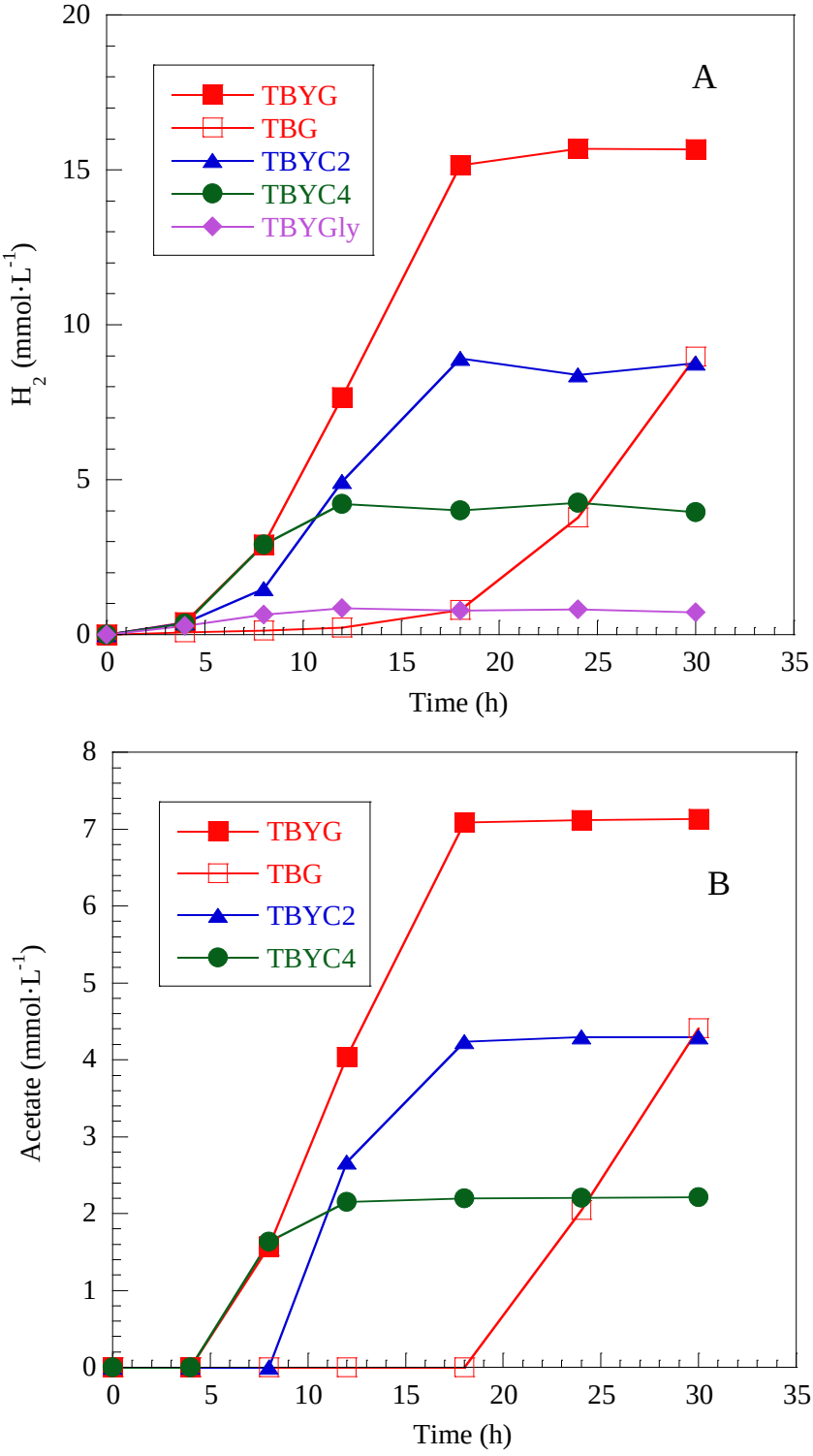


Figure 5.3 H_2 (A) and acetate (B) produced per liquid culture volume as a function of time for different growth medium formulations. TBYGly cultures did not produce acetate. Data points represent averages of duplicate cultures.

5.3.1.1 Glycerol Does Not Sustain Growth

The limited growth and H_2 formation observed in TBYGly cultures was consistent with results obtained previously when cultures were grown on medium with just yeast extract as a carbon and sulfur source (Munro et al. 2009). The static glycerol concentration over the 30 h growth period shown in Figure 5.2 provides further evidence that *T. neapolitana* does not utilize glycerol as a growth substrate.

5.3.1.2 Growth on Glucose, Cellobiose, and Cellotetraose

T. neapolitana has not previously been grown with cellotetraose as a carbon source. *In vitro* evaluation of cellobiose phosphorylase and β-glucosidase activities was conducted with cellotetraose as a substrate and these *T. neapolitana* enzymes were active on cellotetraose as well as cellobiose (Yernool et al. 2000). The batch growth studies presented here reveal that *T. neapolitana* has nearly identical consumption rates for cellobiose and cellotetraose (Figure 5.2). Compared to cultures grown with glucose and yeast extract (TBYG), the cultures with cellobiose and cellotetraose exhibited slower substrate degradation rates, but similar product formation rates.

Two-dimensional protein electrophoresis (2DE) was conducted with TBYG and TBYC4 grown cell samples. 2DE gel images for each condition are presented in Figure 5.4. Only one copy for each condition is shown here, but the same analysis was conducted for biological duplicates. Spots that were further analyzed with MALDI-MS are indicated with red and blue arrows (colors represent selections of independent researchers). The image in Figure 5.4A (the TBYG sample) had increased protein loading compared to the gel in Figure 5.4B. Therefore, the darker

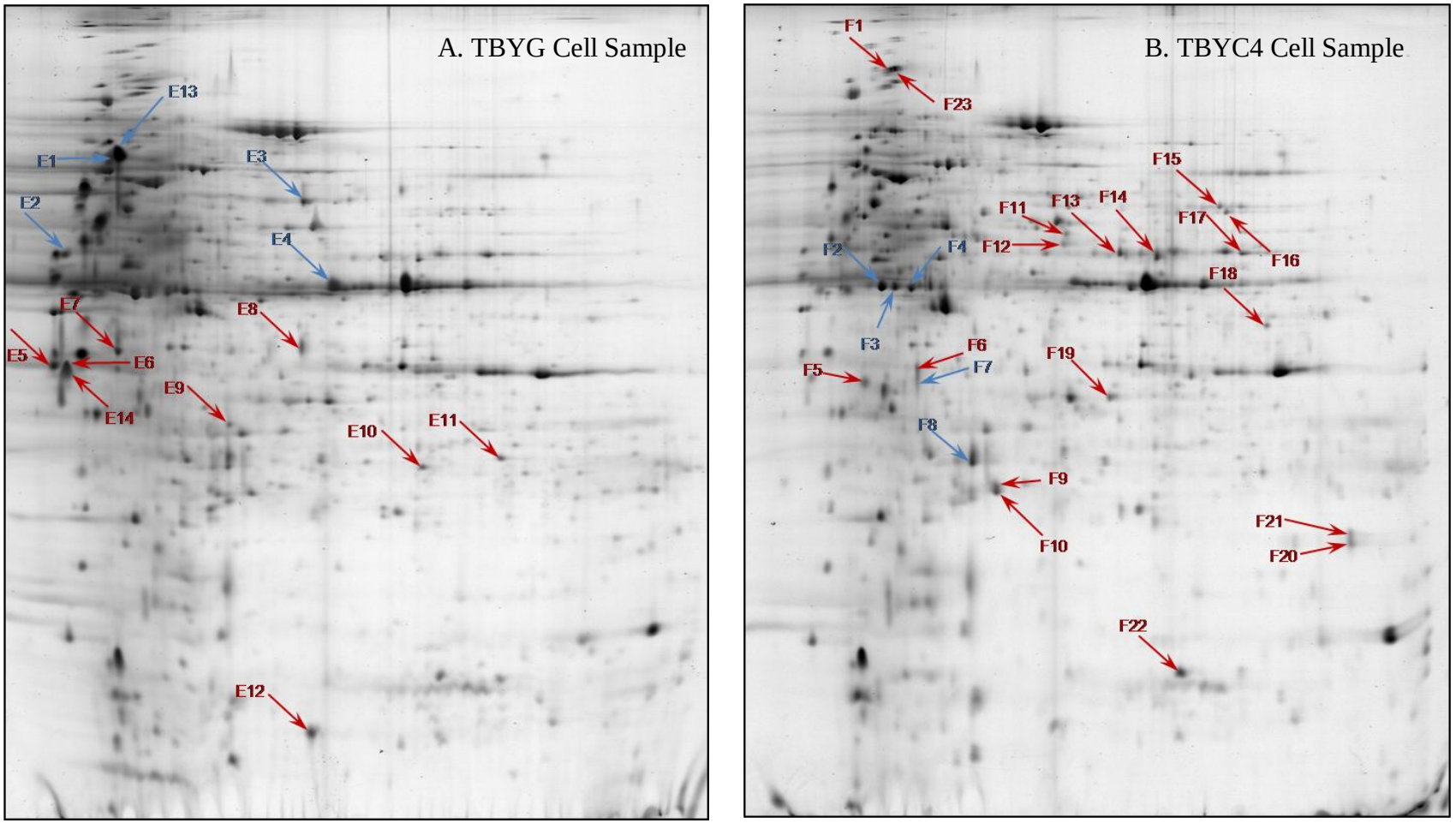


Figure 5.4 TBYG (left, A) and TBYC4 (right, B) 2DE gel images. Red and blue arrows indicate spot selections of two independent researchers. The two gels represent one set of duplicate gels.

spots in Figure 5.4B might indicate increased protein expression levels compared with corresponding spots in Figure 5.4A, since there is increased total protein in Figure 5.4A.

The NCBInr database search results are listed in Table 5.2. Locus tags are included and the relative expression levels between the TBYG and the TBYC4 cells are noted. Several useful results are apparent from this analysis.

An ATP conservation mechanism has previously been proposed based on the expression of cellobiose phosphorylase (CTN_0783) and β-glucosidase (CTN_0782) in *T. neapolitana* (Yernool et al. 2000). In this mechanism, a β-glucosidase cleaves glucose molecules from the ends of cellotetraose, cellotriose, and cellobiose. Cellobiose phosphorylase then cleaves the 1,4-β-linkages of the remaining cellobiose molecules and phosphorylates one of the resulting glucose molecules using inorganic phosphate instead of ATP. The proteomic results show that cellobiose phosphorylase (CTN_0783) was expressed in the cellotetraose grown cells as expected; however, the β-glucosidase (CTN_0782) that was previously identified by Yernool et al. (2000) as active on cellotetraose was not in any of the analyzed spots. This does not mean that this protein is not expressed; it was just not found in this analysis. Instead, endo-1,4-β-glucanase (CTN_1106) had increased expression levels in the TBYC4 cells. This endo-acting cellulase can cleave cellotetraose molecules into two cellobiose molecules.

The presence of endo-1,4-β-glucanase (CTN_1106) and cellobiose phosphorylase (CTN_0783) could indicate the most efficient route of cellotetraose consumption. If endo-1,4-β-glucanase processes all cellotetraose into cellobiose then cellobiose phosphorylase can act on all the resulting cellobiose molecules. Conversely, if β-glucosidase is also acting on cellotetraose molecules, then cleavage

Table 5.2 NCBInr database search results for protein spots picked from the 2DE gels shown in Figure 5.4. The spots were compared between the two gels for expression differences and the following notes were made for each result: relative increase (+), relative decrease (-), low quantity (-,low), and similar results (no change). Best protein scores of 79 or greater are considered statistically significantly ($p<0.05$). No ID indicates that there was no statistically significant hit in the search.

Spot ID	Best Protein Accession	Best Protein Mass	Best Protein Score	Best Protein Description	Locus Tag	TBYG	TBYC4
E1 E6 E13	gi|222099751	37889	390 100 383	Periplasmic binding protein/LacI transcriptional regulator precursor	CTN_0777	+	-
E2	gi|222099563	39253	301	Basic membrane protein	CTN_0589	+	-, low
E3 E9	gi|222099387	101042	477 205	Pyruvate,orthophosphate dikinase	CTN_0413	+	-
E4	gi|222100872	36443	310	Glyceraldehyde-3-phosphate dehydrogenase	CTN_1898	+	-
E5	gi|222099739	43253	436	Maltose ABC transporter, periplasmic maltose-binding protein	CTN_0765	+	-, low
E6 E14	gi|222099634	46525	251 117	Extracellular solute-binding protein, family 1 precursor	CTN_0660	+	-
E7	gi|222099320	38337	450	ABC transporter, periplasmic substrate-binding protein	CTN_0346	+	-
E8				No ID		+	-
E10 E11	gi|24298788	55925	441 329	F-ATPase alpha-subunit	CTN_0846	+	-, low
E12	gi|222099657	22074	224	Pyruvate synthase subunit porC	CTN_0683	+	-
F1 F23	gi|222099757	93620	516 572	Cellobiose-phosphorylase	CTN_0783	-	+
F2 F3 F4	gi|222100871	43782	481 463 577	Bifunctional PGK/TIM	CTN_1896 CTN_1897 CTN_1898	no change	no change
F5	gi|1870180	30173	483	Endo-1,4-β-glucanase	CTN_1106	-, low	+
F6	gi|222099805	24524	194	50S ribosomal protein L25	CTN_0831	-, low	+
F7				No ID		-, low	+
F8	gi|222099827	22570	228	Elongation factor Ts	CTN_0853	-	+

Table 5.2 Continued

F9 F10	gi|222099362	24641	214 234	Transaldolase	CTN_0388	-	+
F11	gi|222099997	50433	555	Transcription termination factor rho	CTN_1023	-	+
F12	gi|222099055	48635	454	Lipopolysaccharide biosynthesis protein	CTN_0081	-, low	+
F13	gi|222100750	44961	636	3-oxoacyl-(Acyl carrier protein) synthase II	CTN_1776	-	+
F14	gi|222099367	45862	79	3-isopropylmalate dehydratase large subunit 1	CTN_0393	-	+
F15	gi|222099451	52523	217	Pyruvate kinase	CTN_0477	-	+
F16	gi|222100218	52590	147	Inosine-5'-monophosphate dehydrogenase	CTN_1244	-	+
F17	gi|222099369	46400	288	6-phosphofructokinase, pyrophosphate-dependent	CTN_0395	-	+
F18	gi|222100270	41632	290	Sugar ABC transporter, ATP-binding protein	CTN_1296		
F19	gi|222099946	32149	362	Dihydrodipicolinate synthase	CTN_0972	-	+
F20 F21	gi|222099093	19382	448 429	Acetolactate synthase, small subunit	CTN_0119	-	+
F22	gi|222099033	15226	431	30S ribosomal protein S6	CTN_0059	-	+

products would also include cellotriose and glucose, both of which cannot be directly processed by cellobiose phosphorylase.

The sugar ABC transporter ATP-binding protein (CTN_1296) was also upregulated in cells grown on cellotetraose compared with glucose, which could indicate that this transport protein is related to growth on cellotetraose. Sequence analysis in the IMG database did not show any other genes related to carbohydrate transport near CTN_1296, so it is unclear if this protein interacts with other proteins for transport activity. Further experimentation is required to identify if this result is significant.

5.3.1.3 Growth on Complex vs. Defined Medium

Comparison of cultures grown with (TBYG) and without yeast extract (TBG) showed a greater glucose consumption rate for cultures grown in the presence of yeast extract. This is partially due to the fact that there was no metabolic adjustment required for the TBYG cultures, since the inoculum had been grown on the same medium. However, in comparison with the cellobiose and cellotetraose grown cultures with YE, the TBG cultures had a significantly longer lag period. This suggests that the cells more quickly adapted to a new carbon source than to a change in sulfur source. After 18 h of incubation the glucose consumption rate in the TBG cultures began to increase and these cultures eventually produced H_2 and acetate at levels comparable to those in the TBYG, TBYC2, and TBYC4 cultures and greater than those achieved in TBYGly grown cultures.

5.3.2 Product and Cell Yields

The product and cell yields for cultures that metabolized glucose, cellobiose, and cellotetraose are shown in Figure 5.5. Acetate yields and H_2 yields were similar between culture conditions. However, the cell yield appears to be significantly higher in the cellotetraose grown cultures compared with the glucose or cellobiose grown cells. Flux balance analysis growth rate predictions from the original *T. maritima* model were also higher on cellotetraose than on glucose or cellobiose (Zhang et al. 2009). This result may be indicative of the proposed ATP conservation mechanism. Further analysis would be required to confirm if cellotetraose grown cells have such an advantage and that the conserved ATP is indeed directed toward biomass formation.

5.3.3 L-Rhamnose Does Not Sustain Growth

Glucose and L-rhamnose were used as growth substrates in a 24 h batch growth experiment. Sugar, acetate, and H_2 concentrations as well as turbidity and total protein were quantified in samples taken at the start and end of the batch growth period. The results are presented in Table 5.3. Although H_2 was formed in the L-rhamnose cultures, as was found with the glycerol cultures, this is likely due to growth on yeast extract. The TBYR cultures also had final protein concentrations similar to TBYGly cultures. Therefore the slight decrease in L-rhamnose concentration could be due to abiotic degradation of L-rhamnose although a specific mechanism is unclear.

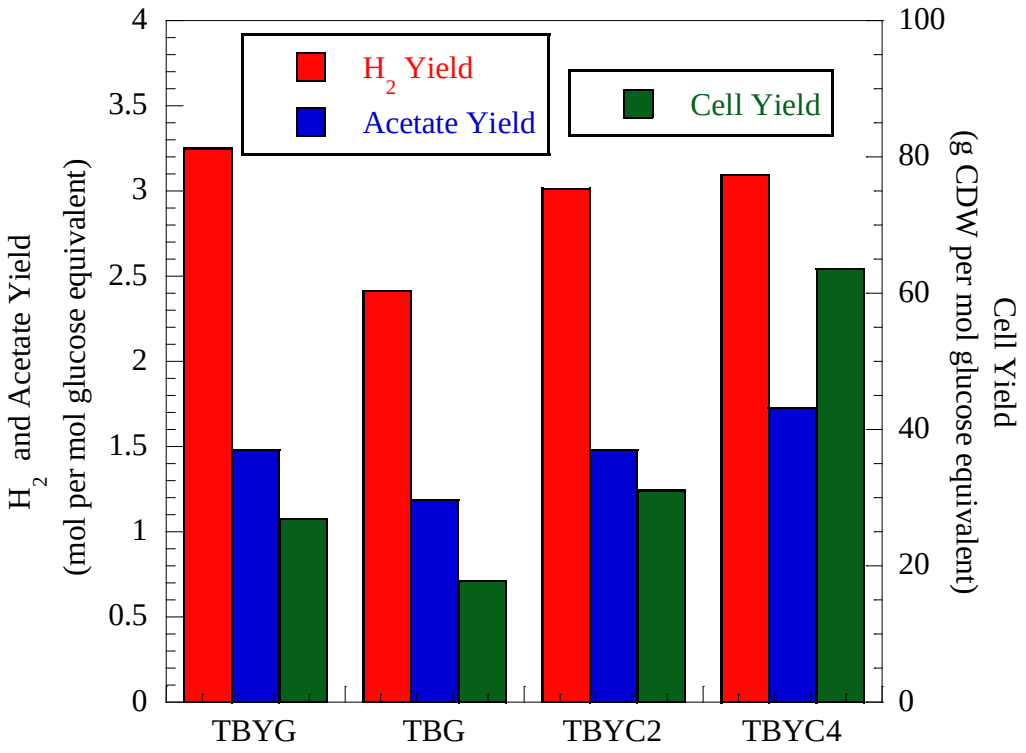


Figure 5.5 H_2 and acetate (left axis) and cell mass (right axis) yields for cultures after 30 h of growth. Each bar represents duplicate culture measurements.

Table 5.3 Results for TBYG and TBYR 24 h batch experiments, standard deviations represent data from triplicate cultures.

	TBYG		TBYR	
	0 h	24 h	0 h	24 h
Substrate Concentration (mmol substrate·L^{-1})	6.0 ± 0.6	0	5.5 ± 0.1	4.4 ± 0.1
H_2 Concentration (mmol·L^{-1})	0	21.1 ± 1.6	0	1.3 ± 0.1
Acetate Concentration (mmol·L^{-1})	0	7.5 ± 0.6	0	0
Protein Concentration (mg Protein·L^{-1})	n.d.[a]	78.2 ± 11.3	n.d.[a]	17.4 ± 1.4

[a] not determined

5.4 Conclusions

In this study, hypotheses related to carbon source degradation in *T. neapolitana* were evaluated. These hypotheses were generated by flux balance analysis of a *T. neapolitana* constraint-based model developed in a previous study (Chapter 4). Flux balance analysis of the model predicted that glycerol, L-rhamnose, and cellotetraose could not support *T. neapolitana* growth.

Batch growth experiments shown here confirmed the *T. neapolitana* model prediction that glycerol and L-rhamnose are not metabolized by *T. neapolitana*. During conversion of the *T. maritima* model to the *T. neapolitana* model, a reaction for glycerol uptake was excluded from the *T. neapolitana* model, because the glycerol uptake facilitator protein (TM1429) present in *T. maritima* was absent from the *T. neapolitana* genome (see Chapter 4, Table 4.2). The negative result for growth on glycerol confirms the validity of excluding this capability from the *T. neapolitana* model. The *T. neapolitana* model did not predict growth on L-rhamnose because four genes associated with L-rhamnose transport in *T. maritima* TM1063 – TM1066 are absent from the *T. neapolitana* genome. The observed lack of growth on L-rhamnose supports the hypothesis that these transport related genes are necessary for growth on L-rhamnose.

The model prediction that cellotetraose could not support growth was disproven by experimental results showing that cellotetraose is consumed by *T. neapolitana* cultures. In the previous analysis of the *T. neapolitana* model compared with the *T. maritima* model (see Chapter 4) it was found that although orthologs for genes associated with cellooligosaccharide transport (TM1219 – TM1222) were found in *T. neapolitana* there was no ortholog for TM1226 which was deemed necessary for the transport reaction by Zhang et al. (2009). This suggests that

either all genes related to growth on cellotetraose have not yet been annotated in the *T. neapolitana* genome or that the TM1226 product is not vital for growth on cellotetraose. Proteomic analysis showed increased expression of a sugar transporter ATP binding protein (CTN_1296, *T. maritima* ortholog TM_1276) in cell samples grown on cellotetraose compared with glucose. Further analysis is required to determine if this protein is associated with cellotetraose transport.

Proteomic analysis of cellotetraose and glucose grown cells also revealed that cellobiose phosphorylase (CTN_0783) and endo-1,4-β-glucanase (CTN_1106) had increased expression in cellotetraose grown cells. Although β-glucosidase (CTN_0782) may still be expressed during growth on cellotetraose it was not identified during proteomic analysis in this study. This suggests that during growth on cellotetraose the ATP conservation mechanism proposed by Yernool et al. (2000) relies on endo-1,4-β-glucanase rather than β-glucosidase. The mechanism may be expanded to include three enzymes, since it is possible that β-glucosidase is still produced. Further analysis would be necessary to determine the expression level of β-glucosidase during growth on various cello-oligomers. Additional proteomic analysis is also underway with cultures grown on glucose with and without yeast extract as a sulfur source. This analysis should allow for identification of the *T. neapolitana* proteins that are expressed for growth on a defined medium with cysteine as a sulfur source. Similar to the work conducted by Roberts et al. (2009), the resulting proteomic data will be applied to the constraint-based model of *T. neapolitana* and should allow the model to successfully simulate growth on cellotetraose and also on a defined medium with cysteine as the sulfur source instead of yeast extract.

Although H_2 and acetate yields were similar for glucose, cellobiose, and cellotetraose grown cultures, the cell growth yield was significantly higher for cellotetraose grown cells compared with the glucose and cellobiose cultures. This

could be evidence of an ATP conservation mechanism that has been proposed for *T. neapolitana* 1,4-β-linked glucose oligomer metabolism (Yernool et al. 2000).

Chapter 6

Conclusions

The goal of this study was to apply the systems biology approach of iterative experimental and computational analysis of biological networks to evaluate *Thermotoga neapolitana* for hydrogen production. The primary metric for a H_2-producing organism is the H_2 yield (mol H_2 per mol glucose). During the first phase of this study the fermentation stoichiometry of *T. neapolitana* was experimentally determined.

Computational modeling of metabolic networks can be used to increase understanding of metabolism and to predict the outcome of metabolic engineering strategies, thereby reducing the need for costly and time-consuming experimental work. In this study, a constraint-based computational model of *T. neapolitana* central carbon metabolism was designed and implemented as a tool to predict carbon source utilization and examine cysteine and hydrogen metabolic pathways. To create the *T. neapolitana* model a novel comparative reconstruction method was demonstrated wherein a constraint-based model for *Thermotoga maritima* was converted into a model for *T. neapolitana* based on synteny between the annotated genome sequences of the two species.

Hypotheses generated from the *T. neapolitana* model were evaluated experimentally in the final phase of this study. Model predictions that glycerol and L-rhamnose were not growth substrates for *T. neapolitana* were validated and the prediction that cellotetraose did not support growth was invalidated. Proteomic analysis of cellotetraose grown cultures was conducted to identify proteins that are

differentially expressed during growth on cellotetraose compared with growth on glucose.

6.1 Revision of Chapter 3: The Fermentation Stoichiometry of *Thermotoga neapolitana* and Influence of Temperature, Oxygen, and pH on Hydrogen Production

Review of Chapter 3 revealed that the hydrogen results were overestimated. An incorrect pressure adjustment was made in the calculation of the mmol of H_2 produced. Thus the hydrogen yield coefficients and volumetric productivity rates were overestimated in the published study. Revisions to the published study will be submitted and the corrected data is presented here in Appendix 1. Changes have been made to Tables 3.1 & 3.2 and Figures 3.2B, 3.4, and 3.5B. The changes impact the reported fermentation stoichiometry (discussed in 6.1.1). The other conclusions in the paper discussed here in section 6.1.2 (Influence of Temperature, O_2, and pH on H_2 Production) are not affected by this correction.

6.1.1 Fermentation Stoichiometry of *T. neapolitana*

Carbon and electron balances on batch cultures of *T. neapolitana* revealed similar fermentation stoichiometry to the related species *T. maritima.* However, the statement published in Munro et al. 2009 that at 85 °C *T. neapolitana* produced 3.8 mol H_2, 1.8 mol acetate, 0.1 mol lactate, and 2 mol CO_2 per mole of glucose consumed was found to be erroneous. The H_2 partial pressure was overestimated based on an incorrect calculation. The correct fermentation stoichiometry at 85 °C for *T. neapolitana* was 2.8 mol H_2, 1.8 mol acetate, 0.1 mol lactate, and 2 mol CO_2 per mole of glucose. This correction is important to note, but does not change the

fundamental result that *T. neapolitana* performs a similar fermentation to that of *T. maritima* and does not exceed the theoretical *in vivo* limit of 4 mol H_2 per mole glucose.

6.1.2 Influence of Temperature, O_2, and pH on H_2 Production

The permissive growth range for *T. neapolitana* has been reported to be 55 – 90 °C with a maximum growth rate achieved at a growth temperature of 77 °C. In this study, five temperatures across the growth range were tested: 60, 65, 70, 77, and 85 °C. Product yields for H_2, acetate, and lactate were evaluated at each temperature and there was no significant difference between growth temperatures. The maximum rate of H_2 production was achieved in cultures grown at 77 and 85 °C. Growth at 77 °C decreases heating costs, growth at 85 °C decreases the risk of contamination by other microbial species, either is sufficient for growth and H_2 production.

The role of O_2 in *T. neapolitana* metabolism had been a debated topic in the literature. Claims of microaerobic metabolism were reported and suggested that H_2 production was increased in the presence of O_2. It has also been noted that *T. neapolitana* has a degree of O_2 tolerance, which is advantageous for industrial fermentation where O_2 exposure can occur.

In this study *T. neapolitana* was exposed to O_2 after 12 h of anaerobic growth and did not exhibit increased H_2 production compared to anaerobic control cultures. This supports other evidence in the literature that *T. neapolitana* does not produce H_2 through microaerobic metabolism to exceed the 4 mol H_2 per mol glucose limit. The result also supports the theory that *T. neapolitana* has some oxygen tolerance given that the production of H_2 continued for several hours following O_2 exposure.

During the temperature variation experiments, there was incomplete removal of fed glucose. To identify the cause of the inhibition, first a range of glucose concentrations were tested to determine if the limited glucose consumption was due to substrate inhibition from the glucose itself. This was not the case, because for low initial glucose concentrations glucose was completely consumed. Next, the effect of pH on glucose consumption was tested. It was found that two methods of increasing pH in *T. neapolitana* cultures, addition of $NaHCO_3$ or an increased initial buffer concentration, allowed the cells to completely consume the available glucose.

6.2 Comparative Constraint-Based Model Reconstruction Method

A novel comparative constraint-based model reconstruction method was demonstrated in this study. In this method an existing constraint-based model for one species can be translated to create a model for a different species, using synteny of the two annotated genomes. This method was demonstrated using the organisms *T. maritima* and *T. neapolitana*, but could be extended for use with other species that are highly related. The resulting models can be compared to identify similarities and differences between the two species.

6.3 *T. neapolitana* Constraint-Based Model

The comparative model reconstruction method was successfully applied to create a constraint-based model of *T. neapolitana* central carbon metabolism using the *T. maritima* constraint-based model as a scaffold. Orthologs for 454 of the 479 genes in the *T. maritima* model were found in the *T. neapolitana* genome and included in the *T. neapolitana* model. Many of the genes absent from the *T. neapolitana* model were

related to oligosaccharide transport and hydrolysis. During model development the non-homologous genes for each species were also analyzed. Several CUT1 family carbohydrate transporters were present in *T. neapolitana* and are absent from *T. maritima*. It is unclear what carbohydrates these proteins might transport.

6.3.1 Flux Balance Analysis (FBA) and Dynamic FBA

Flux balance analysis was used to test the *T. neapolitana* model for growth on 60 carbon sources that had been previously tested with the *T. maritima* model. The *T. neapolitana* model predicted identical growth rates for 44 of the 60 tested carbon sources. The remaining 16 carbon sources did not support *T. neapolitana* growth. These included glycerol, L-rhamnose, mannotriose, mannotetraose and polymers of cellulose (n = 4, 6, and 300 repeat units), galactomannan (n = 4, 6, and 600), glucomannan (n = 4, 6, and 600), and β-1,3/1,4-glucan (n = 4, 6, and 1500). Several of these carbon substrate utilization hypotheses were evaluated with experimental work (Chapter 5 and section 6.5).

Dynamic flux balance analysis (dFBA) was also applied to validate the newly developed *T. neapolitana* model. A H_2 inhibition term was included in the batch growth simulation and the resulting model prediction closely approximated the batch culture experimental data time profiles for biomass, H_2, acetate, and glucose.

6.3.2 Hypothesis Generation: Cysteine and H_2 Metabolism

Constraint-based models can be used to simulate the effects of altering metabolic pathways. The *T. neapolitana* model was used to generate hypotheses related to cysteine and H_2 metabolism. It has been demonstrated experimentally in the

literature and in this study that *T. neapolitana* can grow with cysteine as a sole sulfur source on defined growth medium, whereas *T. maritima* requires either elemental sulfur or a complex sulfur source such as yeast extract. The model was analyzed to evaluate how *T. neapolitana* utilizes cysteine as a sulfur source. Inclusion of a β-cystathionine lyase reaction allowed the model to predict growth on cysteine, however in our analysis of the *T. neapolitana* specific genes (that are not homologous to genes in *T. maritima*) no genes related to this mechanism could be identified. Therefore, this study did not reveal the pathway for cysteine metabolism in *T. neapolitana*.

H_2 metabolism reactions that are relevant to both *T. neapolitana* and *T. maritima* were also probed with the model. Model reactions for H_2-related metabolism were updated to reflect new evidence in the literature regarding a bifurcating hydrogenase that synergistically uses NADH and ferredoxin (Fd) as electron carriers. The inclusion of this reaction in the model resulted in the need for a NADH:Fd oxidoreductase activity to maintain the necessary NADH:Fd ratio required for biomass synthesis. Both proton and sodium translocating oxidoreducatases were tested in the model and simulations predicted growth with the inclusion of either type of oxidoreductase. If the NADH:Fd oxidoreductase is driven by a sodium gradient then additional balancing Na^+- translocating reactions are required in the metabolic network.

6.4 Experimental Evaluation of Model-Derived Hypotheses

Flux balance analysis with the *T. neapolitana* model produced several hypotheses related to carbon substrate utilization. FBA with the model predicted that glycerol and L-rhamnose would not support growth, so each was tested as a carbon source in batch growth. Low levels of growth and hydrogen production in cultures

with glycerol and L-rhamnose were due to metabolism of yeast extract in the growth medium. Glycerol concentrations did not decrease over the 30 h growth period and the slight decrease in L-rhamnose concentration over a 24 h growth period is most likely due to abiotic substrate degradation. Neither substrate supported *T. neapolitana* growth; therefore the model predictions were validated.

FBA of the *T. neapolitana* model also predicted that cellotetraose could not support growth; however batch cultures consumed cellotetraose at rates similar to growth on cellobiose. This result suggests that the cellotetraose hydrolysis and transport reactions in the model need to be revised to reflect the experimental results. Two-dimensional protein electrophoresis (2DE) was conducted to compare cells grown on cellotetraose and glucose to identify differentially expressed proteins. Spots were identified on the 2DE gels and analyzed by mass spectrometry. It was found that during growth on cellotetraose cellobiose phosphorylase and endo-1,4-β-glucanase are expressed, however β-glucosidase was not identified during this analysis (although it may be present). These results suggest that the ATP conservation mechanism for growth on 1,4-β-glucans such as cellotetraose, cellotriose and cellobiose should be amended to include the endo-1,4-β-glucanase along with cellobiose phosphorylase and β-glucanase, especially when the growth substrate is cellotetraose.

6.5 Suggestions for Future Research

Future research could move in several directions. First, the difference in sulfur source metabolism in *T. neapolitana* and *T. maritima* could be explored further. Proteomic analysis is underway to identify differences between cells grown with cysteine as the sole sulfur source with cells grown on yeast extract. This should yield some identifiable protein expression differences between the two growth conditions.

Additional work can also be conducted for proteomic analysis of *T. neapolitana* grown on elemental sulfur compared with cells grown on cysteine, to identify the active metabolic pathways for each sulfur source. An additional valuable result from proteomic analysis of sulfur grown cells would be the identification of the gene transcript associated with the sulfide dehydrogenase.

Following investigation of the sulfur metabolism question, further analysis could be applied to develop the *T. neapolitana* model and to compare *T. neapolitana* and *T. maritima* to gain more insight into how they differ. The *T. neapolitana* model only contains 23% of the total predicted proteins for *T. neapolitana* (26% for *T. maritima*). This means that the function of many proteins in both *T. neapolitana* and *T. maritima* are in need of biochemical verification so a significant amount of potential work is available in this area. In the original *T. maritima* model it is noted that for many metabolites and portions of metabolism there is not enough biochemical information available to connect metabolic pathways with any level of confidence. One example is biotin metabolism. Another example is *Thermotoga* lipids; some knowledge of *T. maritima* lipids is incorporated into the constraint-based models, but lipidomic analysis by LC-MS/MS could be applied to increase understanding of *T. neapolitana* lipids.

An additional research objective that could be explored further is development of a genetic transfer system for *T. neapolitana*. This would significantly improve capacity to test metabolic engineering theories with *T. neapolitana*. Some progress has been made in developing a genetic transfer system using liposome-mediated transfer of the *Thermotoga* sp. RQ2 plasmid (Yu et al. 2001), but the transformation is transient. This is most likely due to the cysteine reducing agent-mediated reduction of chloramphenicol into a nontoxic product (Beckler et al. 1984; Zinder 2008). This could be tested and a more robust system could be developed by testing different

reducing agents. It is also possible to develop new genetic transfer strategies using nanoparticle DNA delivery methods that are not liposome-based and would be compatible with hyperthermophilic species.

Experimental work could also be extended to the development of defined mixed cultures consisting of selected pure cultures in order to increase H_2 yields. Potential organisms that could be tested for co-culture with *T. neapolitana* could include *Clostridium thermocellum* – a crystalline cellulose-degrading H_2-forming species with a preference for hexose over pentose sugars. The use of an acetate-degrading, H_2-producing photofermentative species is also a possibility. Temperature and O_2 requirements will be important to consider in the selection of defined mixed culture member species.

In terms of extending the modeling work demonstrated in this study, constraint-based modeling methods that have been applied here successfully for *T. neapolitana* could also be applied to other species that are relevant to biofuels research and do not already have well developed metabolic models (i.e. *Saccharomyces cerevisiae*). These could include exocellulase producing organisms such as *Thermobifida fusca*, *Clostridia* species, or fungal plant pathogens. The *T. neapolitana* model could also be extended to create a constraint-based model for H_2 producing co-cultures of *T. neapolitana* and *C. thermocellum* or an acetate-consuming photoheterotroph.

APPENDIX 1

Chapter 3 Revisions

Incorrect calculations were used to estimate the quantity of H_2 produced in the work presented here in Chapter 3 and published in Biotechnology Progress (Munro et al. 2009). The conclusions related to growth temperature, O_2 exposure and pH adjustment to relieve inhibition of glucose consumption are still correct with the revised data. However, the hydrogen yield coefficients and volumetric productivity rates were significantly overestimated in the published study and Chapter 3.

In addition to the incorrect calculations, the H_2 calibration curves used for analysis of the original data (peak area from chromatograms obtained by gas chromatography) also required revision. In earlier methods, pure H_2 standards were removed from a H_2 lecture bottle with greater than 1 atm pressure, but were assumed to be at 1 atm. Correct standards have since been prepared by removing pressurized H_2 from the lecture bottle and releasing the pressure lock on the syringe to allow the pressure to reduce to 1 atm. The volume in the syringe was then reduced to the desired amount by depressing the plunger on the syringe. The valve was then closed and used for preparation of standards in serum bottles.

Using the new calibration curve shown here in Figure A1.1, the H_2 data was corrected. The corrected values for Table 3.1 and Table 3.2 from Chapter 3 are shown here in Table 3.1r and Table 3.2r (r indicates "revised"). The revised data is highlighted in yellow in the tables. The corrected figures with hydrogen data are also shown here in Figure 3.2Br, 3.4r, and 3.5Br. The reported fermentation stoichiometry for 85 °C included 3.8 mol H_2 produced per mol of glucose, but the correct yield is 2.8 mol H_2 per mol glucose.

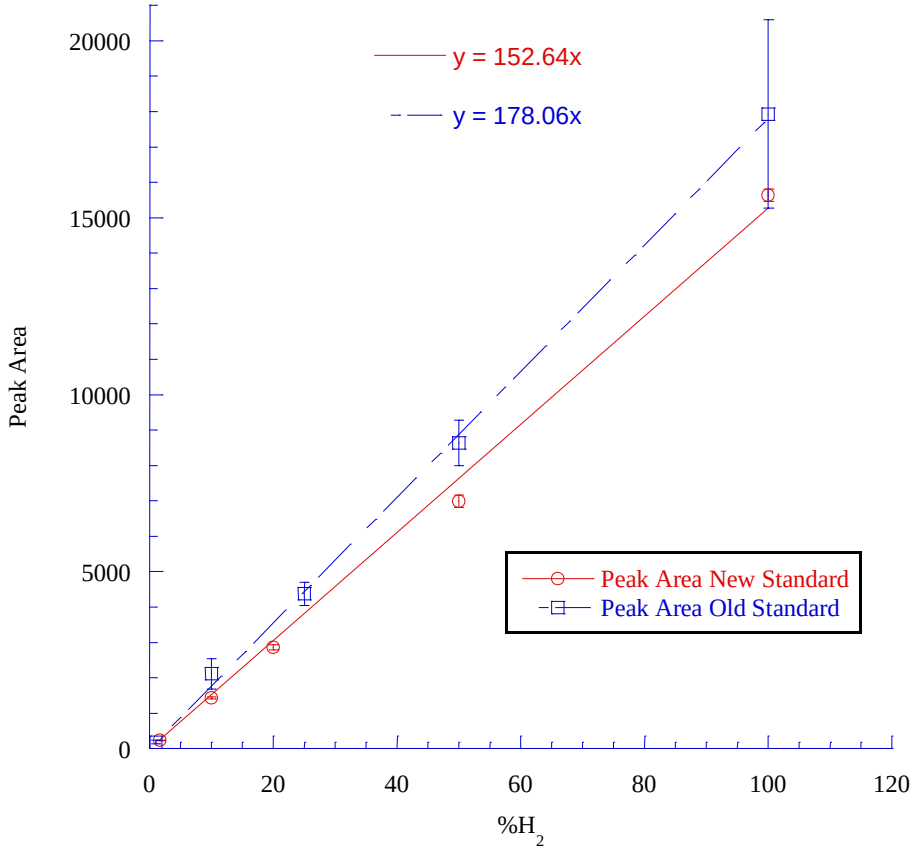


Figure A1.1 (Appendix 1) H_2 Calibration curves used for data in Chapter 3 (Peak Area Old Standard) and the calibration curve used for the corrected data shown here in Appendix 1 (Peak Area New Standard).

Table A1.3.1r (Appendix 1) Revised version of Table 3.1.Yield coefficients of products per mol of glucose consumed and volumetric productivity at various growth temperatures after 48 h of growth. Standard deviations represent culture replicates.

	Temperature				
	60 °C	**65 °C**	**70 °C**	**77 °C**	**85 °C**
Yield Coefficients (mol/ mol)					
$Y_{H_2/Glucose}$	**2.26 ± 0.03**	**2.96 ± 0.29**	**2.77 ± 0.03**	**2.94 ± 0.19**	**2.80 ± 0.28**
$Y_{Acetate/Glucose}$	2.57 ± 0.52	1.62 ± 0.18	1.70 ± 0.06	1.56 ± 0.05	1.87 ± 0.11
$Y_{Lactate/Glucose}$	n.d.[a]	0.019 ± 0.004	0.06 ± 0.02	0.08 ± 0.01	0.12 ± 0.01
$Y_{X/Glucose}$	n.d.	0.37 ± 0.027	0.07 ± 0.037	0.12 ± 0.002	0.09 ± 0.030
Volumetric Productivity ($mmol \cdot L^{-1} \cdot h^{-1}$)					
Q_{H_2}	**0.06 ± 0.009**	**0.26 ± 0.002**	**0.41 ± 0.01**	**0.66 ± 0.03**	**0.62 ± 0.06**
$Q_{Acetate}$	0.06 ±0.004	0.144 ± 0.0003	0.249 ± 0.0004	0.347 ± 0.005	0.40 ± 0.02
$Q_{Lactate}$	n.d.	0.002 ± 0.0001	0.010 ± 0.004	0.018 ± 0.002	0.026 ± 0.001
Q_X	n.d.	0.008 ± 0.0006	0.002 ± 0.00003	0.002 ± 0.0008	0.002 ± 0.0006

[a] not detected

Table A1.3.2r (Appendix 1) Revised version of Table 3.2. Fermentation balance results for 60 h cultures.

Temperature (°C)	Products (mmol/mmol Glucose)					Carbon Balance	Electron Equivalents Balance
	Acetate	Lactate	Biomass[a]	H_2	CO_2		
85	1.84	0.12	0.09	**2.89**	2.35	1.14	**0.99**
77	1.57	0.08	0.12	**3.08**	n.d.[b]	n.d.	**0.92**
70	1.83	0.10	0.06	**2.88**	1.98	1.04	**0.95**
65	1.64	0.02	0.19	**3.09**	1.94	1.04	**0.98**

[a] mmol Biomass was calculated based on the empirical formula $C_5H_8O_3NS_{0.05}$ for *T. maritima*

[b] n.d - not determined

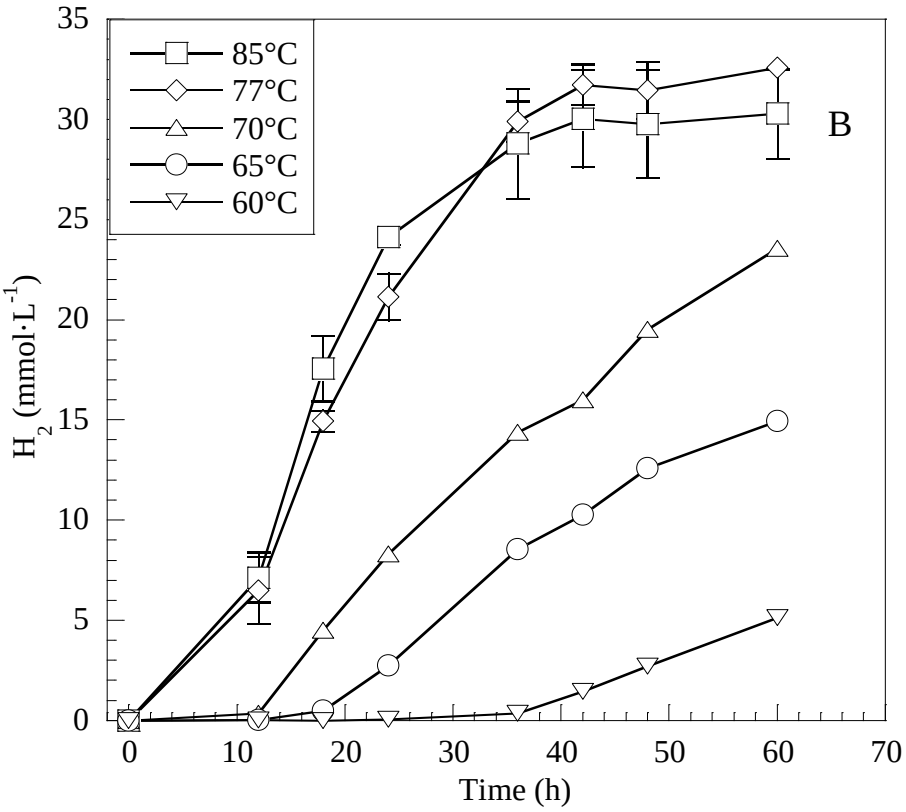


Figure A1.3.2Br (Appendix 1) Revised version of Figure 3.2B. H_2 concentration measurements as a function of time for different temperatures. Error bars represent standard deviation of replicate cultures (n = 3 for 77 °C and n = 4 for 85 °C).

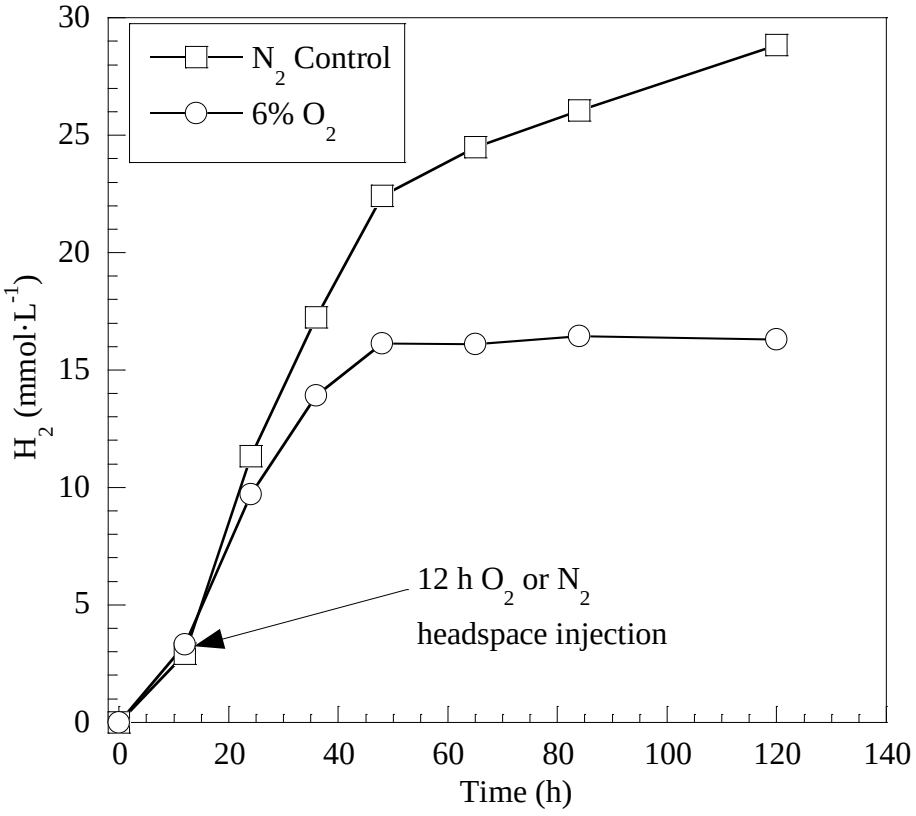


Figure A1.3.4r (Appendix 1) Revised version of Figure 3.4. H_2 concentrations as a function of time for 77 °C cultures with either O_2 or N_2 injections. Data points represent the average of duplicate cultures.

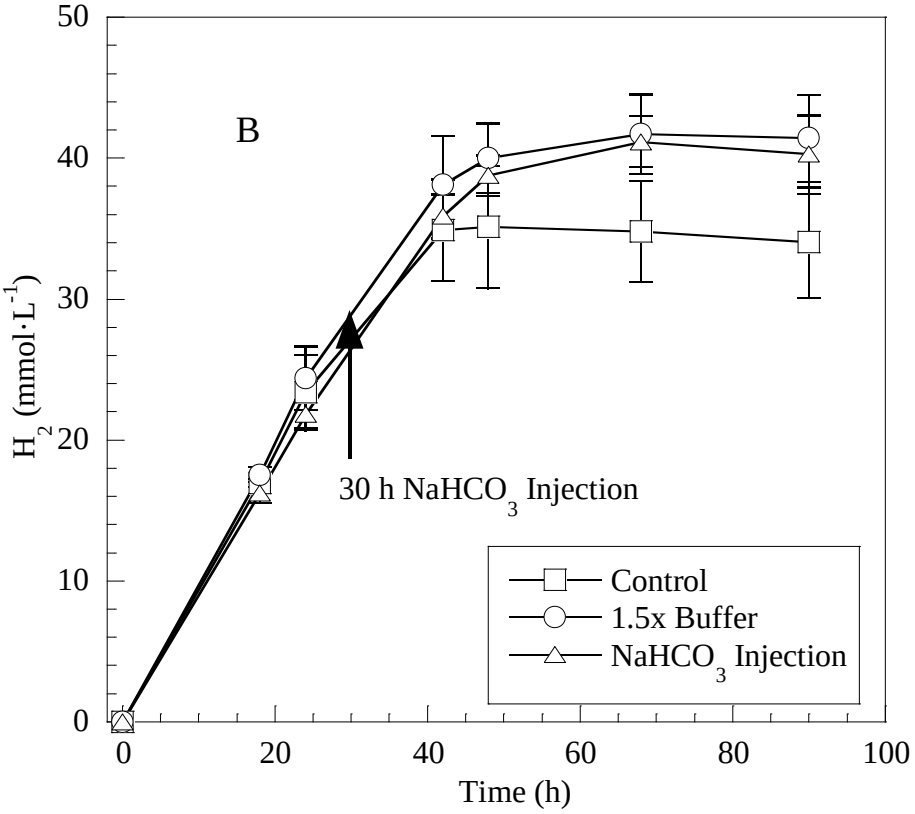


Figure A1.3.5Br (Appendix 1) Revised Version of Figure 3.5B. H_2 concentration in control and pH adjusted cultures with either 1.5x initial buffer concentration or an injection of 1 mL 1M $NaHCO_3$ after 30 h of growth. Error bars represent standard deviation of replicate cultures (n = 4 for each condition).

APPENDIX 2

Species-Specific Gene Lists

The following are lists of non-orthologous genes for *T. maritima* and *T. neapolitana*. These lists were generated using the comparative model reconstruction method detailed in Chapter 4. In the list of *T. maritima* genes, the genes in bold font were used in the *T. maritima* constraint-based model and related reactions were not included in the *T. neapolitana* model due to the absence of orthologs for these genes.

***T. maritima* Gene List**

Genes found in the *T. maritima* genome but not in the *T. neapolitana* genome. Boxes indicate clusters of genes.

***T. maritima* specific genes**

Locus Tag	Product Function	IMG Ortholog neighborhood analysis
TM0001	hypothetical protein	
TM0002	hypothetical protein	
TM0003	hypothetical protein	Genes in **bold font** were in the *T.maritima* model. Finding *T. neapolitana* genes to reproduce the related activities would restore functionality to the model.
TM0004	hypothetical protein	
TM0046	hypothetical protein	
TM0047	transposase putative	
TM0048	transposase	
TM0062	hypothetical protein	
TM0135	transposase	
TM0299	LacI family transcription regulator	
TM0300	**oligopeptide ABC transporter, periplasmic oligopeptide-binding protein, putative, peptide/nickel transport system substrate-binding protein**	Similarity to *Thermotoga* sp. RQ2
TM0301	**oligopeptide ABC transporter, permease protein**	
TM0302	**oligopeptide ABC transporter, permease protein**	
TM0303	**oligopeptide ABC transporter, ATP-binding protein**	
TM0304	**oligopeptide ABC transporter, ATP-binding protein**	
TM0305	**endoglucanase, putative**	
TM0377	hypothetical protein	
TM0390	hypothetical protein	
TM0391	hypothetical protein	
TM0417	hypothetical protein, high-affinity iron transporter	
TM0425	oxidoreductase, putative	
TM0426	PHT4-related protein	Similarity to *Thermotoga petrophila* and *Thermotoga* RQ2
TM0427	**oxidoreductase, putative**	
TM0428	**oxidoreductase, putative**	
TM0429	methyl-accepting chemotaxis protein	
TM0430	sugar ABC transporter, permease	
TM0431	sugar ABC transporter, permease	
TM0432	sugar ABC transporter, periplasmic	

	sugar-binding protein, putative	
TM0433	pectate lyase	
TM0434	**alpha-glucosidase, putative**	TRQ_0505
TM0435	acetyl xylan esterase-related protein	
TM0436	alcohol dehydrogenase, zinc-containing	Similarity to *Thermotoga petrophila*
TM0437	exo-poly-alpha-D-galacturonosidase, putative	
TM0539	tryptophan synthase subunit β, internal gap regions	
TM0559	hypothetical protein	
TM0610	lipopolysaccharide biosynthesis protein	Similarity to *T. petrophila, T. RQ2, T. naphthophila*
TM0611	hypothetical protein	
TM0612	hypothetical protein	
TM0618	hypothetical protein	
TM0620	lipopolysaccharide biosynthesis protein	Similarity to *T. naphthophila*
TM0622	lipopolysaccharide biosynthesis protein, putative	
TM0623	hypothetical protein	
TM0624	N-acetylglucosaminyl-phosphatidylinositol biosynthesis-related protein	
TM0625	hypothetical protein	
TM0626	hypothetical protein	
TM0627	lipopolysaccharide biosynthesis protein	could not find similar group
TM0628	hypothetical protein	
TM0629	hypothetical protein	
TM0630	nucleotide sugar epimerase, putative	
TM0631	lipopolysaccharide biosynthesis protein	
TM0759	acyltransferase, putative	
TM0649	hypothetical protein	
TM0672	hypothetical protein	
TM0776	transposase, putative	
TM0777	transposase	
TM0784	hypothetical protein	
TM1024	hypothetical protein	
TM1025	hypothetical protein	
TM1026	transposase, putative	
TM1044	transposase	
TM1063	**oligopeptide ABC transporter, ATP-binding protein peptide/nickel transport system ATP-binding**	Similarity to *Thermotoga petrophila*

	protein	
TM1064	**oligopeptide ABC transporter, ATP-binding protein, peptide/nickel transport system ATP-binding protein**	
TM1065	**oligopeptide ABC transporter, permease protein peptide/nickel transport system permease protein**	
TM1066	**oligopeptide ABC transporter, permease protein peptide/nickel transport system permease protein**	
TM1120	**glycerol-3-phosphate ABC transporter, periplasmic glycerol-3-phosphate-binding protein**	Similarity to *Thermotoga RQ2, Thermotoga naphthophila, Kosmotoga olearia, T. lettingae, Petrotoga mobilis, Thermosipho* species
TM1121	**glycerol-3-phosphate ABC transporter, permease protein**	
TM1122	**glycerol-3-phosphate ABC transporter, permease protein**	
TM1143	methyl-accepting chemotaxis protein	Not in other *Thermotoga* species, has to do with motility
TM1144	hypothetical protein	
TM1145	hypothetical protein	
TM1146	methyl-accepting chemotaxis protein	
TM1147	hypothetical protein	
TM1173	hypothetical protein	
TM1189	hypothetical protein	
TM1226	**oligopeptide ABC transporter, periplasmic oligopeptide-binding protein, putative**	Present in *T. naphthophila* and *T. RQ2* not in *T. pet, T.nea* or *T. lett*
TM1283	hypothetical protein	Similar to *T. RQ2* and *Fervidobacterium*
TM1284	oxidase-related protein	
TM1285	hypothetical protein	
TM1288	hypothetical protein	Similar to *T. RQ2*
TM1289	ferredoxin	
TM1290	hypothetical protein	
TM1291	iron-sulfur cluster-binding protein	
TM1292	iron-sulfur cluster-binding protein, putative	
TM1323	hypothetical protein	
TM1332	hypothetical protein	
TM1411	helicase-related protein	
TM1412	hypothetical protein	
TM1413	hypothetical protein	

TM1429	**glycerol uptake facilitator protein**	Insertion, not found in *T. pet, T. RQ2, T. nap, T. nea,*it is found in *Carboxydothermus hydrogenoformans, Desulfotomaculum reducens*
TM1589	clostripain-related protein	
TM1613	ATP-synthase F1 delta	
TM1636	hypothetical protein	
TM1669	hypothetical protein	
TM1670	hypothetical protein	
TM1671	hypothetical protein	
TM1677	transposase	
TM1746	**oligopeptide ABC transporter, periplasmic oligopeptide-binding protein**	
TM1747	**oligopeptide ABC transporter, permease protein**	
TM1748	**oligopeptide ABC transporter, permease protein**	
TM1749	**oligopeptide ABC transporter, ATP-binding protein**	Similarity to *Thermotoga RQ2*
TM1750	**oligopeptide ABC transporter, ATP-binding protein**	
TM1751	**endoglucanase**	
TM1752	**endoglucanase**	
TM1753	excinuclease ABC, subunit B-related protein	
TM1757	hypothetical protein	
TM1769	Exodeoxyribonuclease VII small subunit	
TM1779	hypothetical protein	
TM1789	hypothetical protein	
TM1790	hypothetical protein	
TM1791	hypothetical protein	
TM1792	hypothetical protein	
TM1793	hypothetical protein	
TM1794	hypothetical protein	
TM1795	hypothetical protein	Similarity to *Thermotoga RQ2*
TM1796	hypothetical protein	
TM1797	hypothetical protein	
TM1798	hypothetical protein, putative RecB family exonuclease	
TM1799	hypothetical protein	
TM1800	hypothetical protein	
TM1801	hypothetical protein	
TM1802	hypothetical protein	

TM1806	hypothetical protein	
TM1807	hypothetical protein	
TM1808	hypothetical protein	Similarity to *T. RQ2, T. naph, T. pet*
TM1809	hypothetical protein	
TM1810	hypothetical protein	
TM1811	hypothetical protein	
TM1813	hypothetical protein	
TM1829	hypothetical protein	
TM1830	hypothetical protein	
TM1831	transposase, putative	
TM1832	transposase	
TM1833	methyl-accepting chemotaxis-related protein	

***T. neapolitana* Gene List**

Genes found in the *T. neapolitana* genome but not in the *T. maritima* genome. Boxes indicate clusters of genes.

***T. neapolitana* specific genes**

Locus Tag	Product Function	IMG Ortholog neighborhood analysis
CTN_0026	hypothetical protein	
CTN_0027	sugar transferase	
CTN_0028	glycosyl transferase, group 1	
CTN_0029	O-antigen polymerase	
CTN_0030	glycosyl transferase, family 2	
CTN_0031	glucose-1-phosphate thymidylyltransferase	Similarity to *Thermotoga* sp. RQ2: NC_010483
CTN_0032	putative dTDP-6-deoxy-D-glucose-3,5-epimerase	
CTN_0033	dTDP-glucose 4,6-dehydratase	
CTN_0034	hypothetical protein	
CTN_0035	putative dTDP-4-dehydrorhamnose reductase	
CTN_0036	putative repeat unit transporter	
CTN_0037	hypothetical protein	
CTN_0045	hypothetical protein	
CTN_0046	hypothetical protein	
CTN_0047	ATPase-like protein	
CTN_0048	hypothetical protein	
CTN_0049	chromosome segregation ATPase-like protein precursor	
CTN_0050	metal dependent phosphohydrolase	
CTN_0086	CRISPR related	
CTN_0105	Methyltransferase type 11	
CTN_0109	hypothetical protein	
CTN_0130	transposase	
CTN_0235	TRAP dicarboxylate transporter, DctM subunit	TnapDRAFT_0561
CTN_0236	tripartite ATP-independent periplasmic transporter, DctQ component precursor	TnapDRAFT_0560
CTN_0237	C4-dicarboxylate-binding protein	TnapDRAFT_0559
CTN_0238	ABC transporter related	TnapDRAFT_0558
CTN_0239	inner-membrane translocator	TnapDRAFT_0557
CTN_0240	sugar binding protein of ABC transporter	TnapDRAFT_0556

CTN_0241	β-lactamase domain protein	similar to Tlet_1093
CTN_0242	ABC-type, ATP binding, spermidine/spermine transporter, PotA	similar to *Marinomonas, Ochrobactrum, Klebsiella*
CTN_0243	binding-protein-dependent transport systems inner membrane component precursor	
CTN_0244	binding-protein-dependent transport systems inner membrane component precursor	
CTN_0245	extracellular solute-binding protein family 1 precursor	
CTN_0289	biotin synthase	identical to *Thermotoga petrophila*
CTN_0319	CRISPR related	
CTN_0355	hypothetical protein	**No Thermotoga homology** - *Clostridium bolteae, Clavibacter?Bacillus licheniformis, Catenulispora, B. licheniformis, Bacillus clausii*
CTN_0356	putative sugar ABC transporter, permease component	
CTN_0357	hypothetical protein	
CTN_0358	extracellular solute-binding protein family 1 precursor	
CTN_0359	ureidoglycolate hydrolase	
CTN_0360	monosaccharide-transporting ATPase precursor	
CTN_0361	ABC transporter related	
CTN_0362	ribose ABC transporter, permease protein	
CTN_0363	hypothetical protein	
CTN_0364	putative periplasmic binding protein	
CTN_0365	oxidoreductase, short chain dehydrogenase/reductase family	
CTN_0366	putative ribose/galactose/methyl galactoside import ATP-binding protein	
CTN_0367	monosaccharide-transporting ATPase	Tlet_0185
CTN_0368	hypothetical protein	
CTN_0369	alcohol dehydrogenase GroES domain protein	
CTN_0370	hypothetical protein	
CTN_0371	transcriptional regulator, GntR family	Tlet_0191
CTN_0372	betaine-aldehyde dehydrogenase	Tlet_0192
CTN_0373	glucose-6-phosphate isomerase	Tlet_0193
CTN_0383	fructose-bisphosphate aldolase	similar to TnapDRAFT_1627,Tpet_0612
CTN_0384	carbohydrate kinase, FGGY	similar to TnapDRAFT_1626, Tpet_0613

CTN_0545	hypothetical protein	
CTN_0555	β-lactamase domain protein	
CTN_0617	endo-1,3-β-xylanase	
CTN_0633	transposase	
CTN_0660	extracellular solute-binding protein, family 1 precursor	Similar to *Kosmotoga olearia, Thermosipho, Petrotoga*
CTN_0661	binding-protein-dependent transport systems inner membrane component precursor	
CTN_0662	binding-protein-dependent transport systems inner membrane component	
CTN_0673	NADPH-dependent FMN reductase	Insertion no match to other Thermotoga species might be a misannotation
CTN_0674	hypothetical protein	
CTN_0690	hypothetical protein	
CTN_0695	hypothetical protein	Very small sequences, no apparent homology to other *Thermotoga* species, might be CRISPR related
CTN_0696	ganglioside-induced differentiation-associated protein 1-like 1	
CTN_0697	protein containing DUF1232	
CTN_0698	hypothetical protein	
CTN_0699	hypothetical protein	
CTN_0706	hypothetical protein	homologous to *Thermotoga petrophila* CRISPR related proteins
CTN_0707	CRISPR-associated protein, Cas6 family	
CTN_0708	hypothetical protein	
CTN_0709	CRISPR-associated autoregulator, Cst2 family	
CTN_0710	CRISPR-associated protein Cas5 family	
CTN_0711	CRISPR-associated helicase, Cas3 family	
CTN_0712	CRISPR-associated helicase, Cas3 family	
CTN_0713	CRISPR-associated exonuclease, Cas4 family	
CTN_0714	CRISPR-associated protein, Cas1 family	
CTN_0715	CRISPR-associated protein, Cas2 family	
CTN_0716	hypothetical membrane spanning protein	
CTN_0717	hypothetical protein	
CTN_0746	hypothetical protein	
CTN_0774	ROK family protein	TRQ2_0976
CTN_0775	ABC transporter related	Similar to TRQ2_0975

CTN_0776	monosaccharide-transporting ATPase	Similar to TRQ2_0974
CTN_0777	periplasmic binding protein/LacI transcriptional regulator precursor	Similar to TRQ2_0973
CTN_0778	binding-protein-dependent transport systems inner membrane component precursor	Similar to TRQ2_0972
CTN_0779	binding-protein-dependent transport systems inner membrane component	Similar to TRQ2_0971
CTN_0780	extracellular solute-binding protein, family 1 precursor	Similar to TRQ2_0970
CTN_0781	alpha amylase, catalytic region precursor	Similar to Tpet_0953 and TnapDRAFT_1683
CTN_0782	β-glucosidase A	Similar to Tpet_0952 and TnapDRAFT_1684
CTN_0822	DNA break repair rad50 ATPase	
CTN_0845	F-ATPase delta subunit	
CTN_0913	metallophosphoesterase	
CTN_0914	hypothetical protein	
CTN_0915	H+transporting two-sector ATPase C (AC39) subunit	Similar to TRQ genes 1109 - 1100, slightly different annotation, also similar to *Halothermothrix orenii*, and *Clostridium tetani* **-I've tested Na+ driven V-ATPase activity in the model**
CTN_0916	putative A-ATPase I-subunit	
CTN_0917	V-ATPase F-subunit	
CTN_0918	V-ATPase G-subunit	
CTN_0919	V-ATPase E-subunit	
CTN_0920	V-ATPase A-subunit	
CTN_0921	V-ATPase B-subunit	
CTN_0922	V-ATPase D-subunit	
CTN_0959	Exodeoxyribonuclease 7 small subunit	
CTN_1080	hypothetical protein	*Ralstonia*
CTN_1081	hypothetical protein	*Fervidobacterium, Thermosipho*
CTN_1082	lipopolysaccharide biosynthesis protein-related protein	TRQ2_1419
CTN_1092	Putative metal dependent phosphohydrolase precursor	
CTN_1103	hypothetical protein	partial homology to TM1322 & TM1332
CTN_1104	hypothetical protein	No direct match in any *Thermotoga* species, might be CRISPR repeat associated
CTN_1105	hypothetical protein	
CTN_1172	ribonuclease H	
CTN_1173	hypothetical protein	no clear matches
CTN_1174	hypothetical protein	
CTN_1175	thioredoxin	
CTN_1185	hypothetical protein	

CTN_1186	hypothetical protein	
CTN_1251	hypothetical protein	
CTN_1262	hypothetical protein	
CTN_1263	putative uncharacterized protein precursor	Similar to Tpet and TnapDRAFT sequences
CTN_1264	ABC transporter related precursor	
CTN_1265	Fe-S oxidoreductase-like protein	
CTN_1266	hypothetical protein	Similar to Tpet_1462 some overlap with TM1322 and TM1332 inserted between
CTN_1302	endo-1,4-β-xylanase precursor	Matches TRQ2_1549, similar to *Clostridium papyrosolvens* and *cellulolyticum*
CTN_1372	extracellular solute-binding protein family 1	Matches with *Fervidobacterium nodosum, Thermotoga lettingae, Halothermothrix orenii*
CTN_1373	binding-protein-dependent transport systems inner membrane component	
CTN_1374	binding-protein-dependent transport systems inner membrane component	
CTN_1386	diguanylate cyclase	Annotated as a transposase in other *Thermotoga* species (Tnap. TRQ. Tpet)
CTN_1387	diguanylate cyclase	In Tnap draft
CTN_1504	hypothetical protein	
CTN_1539	ABC transporter related precursor	partial homology withTM1028
CTN_1540	sugar:cation symporter family protein	No full sequence homologues
CTN_1541	extracellular solute-binding protein, family 1 precursor	
CTN_1542	binding-protein-dependent transport systems inner membrane component	
CTN_1543	binding-protein-dependent transport systems inner membrane component	
CTN_1544	fructose-6-phosphate aldolase	
CTN_1545	hypothetical protein	
CTN_1546	hypothetical protein	
CTN_1547	hypothetical protein	
CTN_1548	aldehyde dehydrogenase	
CTN_1549	hypothetical protein	
CTN_1550	alpha-glucosidase	
CTN_1551	transcriptional regulator	
CTN_1552	hypothetical protein	
CTN_1553	hypothetical protein	
CTN_1554	methyl-accepting chemotaxis sensory transducer precursor	

CTN_1555	putative uncharacterized protein precursor	partial homology with TM1027
CTN_1794	radical SAM domain protein	*Bacillus cereus* SAM protein
CTN_1795	radical SAM domain protein	
CTN_1796	hypothetical protein	some overlap with TM0784
CTN_1813	transposase	
CTN_1820	hypothetical protein	
CTN_1821	hypothetical protein	
CTN_1914	hypothetical protein	
CTN_1930	putative uncharacterized protein precursor	*Thermosipho africanus* similarity
CTN_1931	putative uncharacterized protein precursor	
CTN_1932	putative uncharacterized protein precursor	
CTN_1933	secretin/TonB, short N-terminal domain precursor	
CTN_1934	secretin/TonB, short N-terminal domain precursor	
CTN_1935	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase	
CTN_1936	hypothetical protein	
CTN_1937	transcriptional regulator, SARP family	

REFERENCES

Aceves-Lara, C. A., E. Latrille, N. Bernet, P. Buffiere and J. P. Steyer (2008). "A pseudo-stoichiometric dynamic model of anaerobic hydrogen production from molasses." Water Research **42**(10-11): 2539-2550.

Adams, M. W. W. (1990). "The metabolism of hydrogen by extremely thermophilic, sulfur-dependent bacteria." Fems Microbiology Reviews **75**(2-3): 219-237.

Akkerman, I., M. Janssen, J. Rocha and R. H. Wijffels (2002). "Photobiological hydrogen production: photochemical efficiency and bioreactor design." International Journal of Hydrogen Energy **27**(11-12): 1195-1208.

Amend, J. P. and E. L. Shock (2001). "Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria." Fems Microbiology Reviews **25**(2): 175-243.

Andrews, J. F. (1968). "A mathematical model for continuous culture of microorganisms utilizing inhibitory substrates." Biotechnology and Bioengineering **10**(6): 707-723.

Angenent, L. T., K. Karim, M. H. Al-Dahhan and R. Domiguez-Espinosa (2004). "Production of bioenergy and biochemicals from industrial and agricultural wastewater." Trends in Biotechnology **22**(9): 477-485.

Backiel, J., O. Juárez, D. V. Zagorevski, Z. Wang, M. J. Nilges and B. Barquera (2008). "Covalent Binding of Flavins to RnfG and RnfD in the Rnf Complex from *Vibrio cholerae*." Biochemistry **47**(43): 11273-11284.

Bailey, J. E. and D. F. Ollis (1986). Biochemical engineering fundamentals. New York, McGraw-Hill.

Beard, D. A., S. D. Liang and H. Qian (2002). "Energy balance for analysis of complex metabolic networks." Biophys J **83**(1): 79-86.

Becker, S. A., A. M. Feist, M. L. Mo, G. Hannum, B. O. Palsson and M. J. Herrgard (2007). "Quantitative prediction of cellular metabolism with constraint-based models: The COBRA toolbox." Nature Protocols **2**(3): 727-738.

Beckler, G. S., L. A. Hook and J. N. Reeve (1984). "Chloramphenicol acetyltransferase should not provide methanogens with resistance to chloramphenicol." Applied and Environmental Microbiology **47**(4): 868-869.

Belkin, S., C. O. Wirsen and H. W. Jannasch (1986). "A new sulfur-reducing, extremely thermophilic eubacterium from a submarine thermal vent." Applied and Environmental Microbiology **51**(6): 1180-1185.

Benemann, J. R. (1997). "Feasibility analysis of photobiological hydrogen production." International Journal of Hydrogen Energy **22**(10-11): 979-987.

Berezina, O. V., N. A. Lunina, V. V. Zverlov, D. G. Naumoff, W. Liebl and G. A. Velikodvorskaya (2003). "A cluster of *Thermotoga neapolitana* genes involved in the degradation of starch and maltodextrins: The molecular structure of the locus." Molecular Biology **37**(5): 678-685.

Berezina, O. V., V. V. Zverlov, N. A. Lunina, L. A. Chekanovskaya, E. N. Dubinina, W. Liebl and G. A. Velikodvorskaya (1999). "Gene and properties of thermostable 4-α-glucanotransferase of *Thermotoga neapolitana*." Molecular Biology **33**(5): 801-806.

Bernard, O. and G. Bastin (2005). "Identification of reaction networks for bioprocesses: Determination of a partially unknown pseudo-stoichiometric matrix." Bioprocess and Biosystems Engineering **27**(5): 293-301.

Bernard, O. and G. Bastin (2005). "On the estimation of the pseudo-stoichiometric matrix for macroscopic mass balance modelling of biotechnological processes." Mathematical Biosciences **193**(1): 51-77.

Bibel, M., C. Brettl, U. Gosslar, G. Kriegshauser and W. Liebl (1998). "Isolation and analysis of genes for amylolytic enzymes of the hyperthermophilic bacterium *Thermotoga maritima*." Fems Microbiology Letters **158**(1): 9-15.

Blomberg, A. (2004). Proteomics in metabolic engineering. Metabolic Engineering in the Post-Genomic Era. B. N. Kholodenko and H. V. Westerhoff. Wymondham, UK, Horizon Bioscience**:** 37-68.

Boiangiu, C. D., E. Jayamani, D. Brügel, G. Herrmann, J. Kim, L. Forzi, R. Hedderich, I. Vgenopoulou, A. J. Pierik, J. Steuber and W. Buckel (2005). "Sodium ion pumps and hydrogen production in glutamate fermenting anaerobic bacteria." Journal of Molecular Microbiology and Biotechnology **10**(2-4): 105-119.

Bok, J. D., D. A. Yernool and D. E. Eveleigh (1998). "Purification, characterization, and molecular analysis of thermostable cellulases CelA and CelB from *Thermotoga neapolitana*." Applied and Environmental Microbiology **64**(12): 4774-4781.

Borodin, V. B., A. A. Tsygankov, K. K. Rao and D. O. Hall (2000). "Hydrogen production by *Anabaena variabilis* PK84 under simulated outdoor conditions." Biotechnology and Bioengineering **69**(5): 478-485.

Box, G. E. P. and W. J. Hill (1967). "Discrimination among mechanistic models." Technometrics **9**(1): 57-71.

Boyle, N. R. and J. A. Morgan (2009). "Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*." BMC Syst Biol **3**: 4.

Bruggemann, H., S. Baumer, W. F. Fricke, A. Wiezer, H. Liesegang, I. Decker, C. Herzberg, R. Martinez-Arias, R. Merkl, A. Henne and G. Gottschalk (2003). "The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease." Proceedings of the National Academy of Sciences of the United States of America **100**(3): 1316-1321.

Chen, C. C., C. Y. Lin and J. S. Chang (2001). "Kinetics of hydrogen production with continuous anaerobic cultures utilizing sucrose as the limiting substrate." Applied Microbiology and Biotechnology **57**(1-2): 56-64.

Chen, W. H., S. Y. Chen, S. K. Khanal and S. W. Sung (2006). "Kinetic study of biological hydrogen production by anaerobic fermentation." International Journal of Hydrogen Energy **31**(15): 2170-2178.

Chhabra, S. R., K. R. Shockley, S. B. Conners, K. L. Scott, R. D. Wolfinger and R. M. Kelly (2003). "Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*." Journal of Biological Chemistry **278**(9): 7540-7552.

Childers, S. E. and K. M. Noll (1994). "Characterization and regulation of sulfur reductase-activity in *Thermotoga neapolitana*." Applied and Environmental Microbiology **60**(7): 2622-2626.

Childers, S. E., M. Vargas and K. M. Noll (1992). "Improved methods for cultivation of the extremely thermophilic bacterium *Thermotoga neapolitana*." Applied and Environmental Microbiology **58**(12): 3949-3953.

Chin, H. L., Z. S. Chen and C. P. Chou (2003). "Fedbatch operation using *Clostridium acetobutylicum* suspension culture as biocatalyst for enhancing hydrogen production." Biotechnology Progress **19**(2): 383-388.

Chou, C. J., K. R. Shockley, S. B. Conners, D. L. Lewis, D. A. Comfort, M. W. W. Adams and R. M. Kelly (2007). "Impact of substrate glycoside linkage and elemental sulfur on bioenergetics of and hydrogen production by the hyperthermophilic Archaeon *Pyrococcus furiosus*." Applied and Environmental Microbiology **73**(21): 6842-6853.

Claassen, P. A. M., J. B. van Lier, A. M. L. Contreras, E. W. J. van Niel, L. Sijtsma, A. J. M. Stams, S. S. de Vries and R. A. Weusthuis (1999). "Utilisation of biomass for the supply of energy carriers." Applied Microbiology and Biotechnology **52**(6): 741-755.

Conners, S. B., E. F. Mongodin, M. R. Johnson, C. I. Montero, K. E. Nelson and R. M. Kelly (2006). "Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species." Fems Microbiology Reviews **30**(6): 872-905.

Conners, S. B., C. I. Montero, D. A. Comfort, K. R. Shockley, M. R. Johnson, S. R. Chhabra and R. M. Kelly (2005). "An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*." Journal of Bacteriology **187**(21): 7267-7282.

Covert, M. W., C. H. Schilling and B. Palsson (2001). "Regulation of gene expression in flux balance models of metabolism." Journal of Theoretical Biology **213**(1): 73-88.

Darling, A. C. E., B. Mau, F. R. Blattner and N. T. Perna (2004). "Mauve: Multiple alignment of conserved genomic sequence with rearrangements." Genome Research **14**(7): 1394-1403.

Das, D. and T. N. Veziroglu (2001). "Hydrogen production by biological processes: a survey of literature." International Journal of Hydrogen Energy **26**(1): 13-28.

Datar, R., J. Huang, P. C. Maness, A. Mohagheghi, S. Czemik and E. Chornet (2007). "Hydrogen production from the fermentation of corn stover biomass pretreated with a steam-explosion process." International Journal of Hydrogen Energy **32**(8): 932-939.

de Vrije, T., R. R. Bakker, M. A. Budde, M. H. Lai, A. E. Mars and P. A. Claassen (2009). "Efficient hydrogen production from the lignocellulosic energy crop *Miscanthus* by the extreme thermophilic bacteria *Caldicellulosiruptor saccharolyticus* and *Thermotoga neapolitana*." Biotechnol Biofuels **2**(1): 12.

de Vrije, T., G. G. de Haas, G. B. Tan, E. R. P. Keijsers and P. A. M. Claassen (2002). "Pretreatment of *Miscanthus* for hydrogen production by *Thermotoga elfii*." International Journal of Hydrogen Energy **27**(11-12): 1381-1390.

DeBoy, R. T., E. F. Mongodin, J. B. Emerson and K. E. Nelson (2006). "Chromosome evolution in the *Thermotogales*: Large-scale inversions and strain diversification of CRISPR sequences." Journal of Bacteriology **188**(7): 2364-2374.

Demain, A. L., M. Newcomb and J. H. D. Wu (2005). "Cellulase, Clostridia, and ethanol." Microbiology and Molecular Biology Reviews **69**(1): 124-154.

Dharmadi, Y. and R. Gonzalez (2004). "DNA microarrays: Experimental issues, data analysis, and application to bacterial systems." Biotechnology Progress **20**(5): 1309-1324.

Dietrich, G., N. Weiss and J. Winter (1988). "*Acetothermus Paucivorans*, gen-nov, sp-nov, a strictly anaerobic, thermophilic bacterium from sewage-sludge, fermenting hexoses to acetate, CO_2 and H_2." Systematic and Applied Microbiology **10**(2): 174-179.

Duffaud, G. D., C. M. McCutchen, P. Leduc, K. N. Parker and R. M. Kelly (1997). "Purification and characterization of extremely thermostable β-mannanase, β-mannosidase, and α-galactosidase from the hyperthermophilic eubacterium *Thermotoga neapolitana* 5068." Applied and Environmental Microbiology **63**(1): 169-177.

Edwards, J. S., M. Covert and B. Palsson (2002). "Metabolic modelling of microbes: The flux-balance approach." Environmental Microbiology **4**(3): 133-140.

Edwards, J. S. and B. O. Palsson (1999). "Systems properties of the *Haemophilus influenzae* Rd metabolic genotype." Journal of Biological Chemistry **274**(25): 17410-17416.

Edwards, J. S. and B. O. Palsson (2000). "The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities." Proceedings of the National Academy of Sciences of the United States of America **97**(10): 5528-5533.

Energy, U. D. o. (2008). Research solicitation: Systems biology, model organism development, and enzyme discovery for biological hydrogen production, US Department of Energy.

Eriksen, N. T., T. M. Nielsen and N. Iversen (2008). "Hydrogen production in anaerobic and microaerobic *Thermotoga neapolitana*." Biotechnology Letters **30**(1): 103-109.

Fang, H. H. P. and H. Liu (2002). "Effect of pH on hydrogen production from glucose by a mixed culture." Bioresource Technology **82**(1): 87-93.

Fang, H. H. P., T. Zhang and H. Liu (2002). "Microbial diversity of a mesophilic hydrogen-producing sludge." Applied Microbiology and Biotechnology **58**(1): 112-118.

Feist, A. M., C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis and B. O. Palsson (2007). "A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." Mol Syst Biol **3**: 121.

Feist, A. M., J. C. Scholten, B. O. Palsson, F. J. Brockman and T. Ideker (2006). "Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*." Mol Syst Biol **2**: 2006 0004.

Finehout, E. J. and K. H. Lee (2003). "Comparison of automated in-gel digest methods for femtomole level samples." Electrophoresis **24**(19-20): 3508-3516.

Flynn, T., M. L. Ghirardi and M. Seibert (2002). "Accumulation of O_2-tolerant phenotypes in H_2-producing strains of *Chlamydomonas reinhardtii* by sequential applications of chemical mutagenesis and selection." International Journal of Hydrogen Energy **27**(11-12): 1421-1430.

Forster, J., I. Famili, P. Fu, B. O. Palsson and J. Nielsen (2003). "Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network." Genome Research **13**(2): 244-253.

Galperin, M. Y., K. M. Noll and A. H. Romano (1996). "The glucose transport system of the hyperthermophilic anaerobic bacterium *Thermotoga neapolitana*." Applied and Environmental Microbiology **62**(8): 2915-2918.

Galperin, M. Y., K. M. Noll and A. H. Romano (1997). "Coregulation of β-galactoside uptake and hydrolysis by the hyperthermophilic bacterium *Thermotoga neapolitana*." Applied and Environmental Microbiology **63**(3): 969-972.

Garcia, D. E., E. E. Baidoo, P. I. Benke, F. Pingitore, Y. J. Tang, S. Villa and J. D. Keasling (2008). "Separation and mass spectrometry in microbial metabolomics." Current Opinion in Microbiology **11**(3): 233-239.

Gottschalk, G. (1986). Bacterial metabolism. New York, Springer-Verlag.

Haldane, J. B. S. (1930). Enzymes. London, New York,, Longmans, Green.

Hallenbeck, P. C. (2005). "Fundamentals of the fermentative production of hydrogen." Water Science and Technology **52**(1-2): 21-29.

Hallenbeck, P. C. and J. R. Benemann (2002). "Biological hydrogen production: Fundamentals and limiting processes." International Journal of Hydrogen Energy **27**(11-12): 1185-1193.

Han, K. and O. Levenspiel (1988). "Extended Monod kinetics for substrate, product, and cell-inhibition." Biotechnology and Bioengineering **32**(4): 430-437.

Hansel, A. and P. Lindblad (1998). "Towards optimization of Cyanobacteria as biotechnologically relevant producers of molecular hydrogen, a clean and renewable energy source." Applied Microbiology and Biotechnology **50**(2): 153-160.

Hansen, T. and P. Schonheit (2003). "ATP-dependent glucokinase from the hyperthermophilic bacterium *Thermotoga maritima* represents an extremely thermophilic ROK glucokinase with high substrate specificity." Fems Microbiology Letters **226**(2): 405-411.

Hatzimanikatis, V., L. H. Choe and K. H. Lee (1999). "Proteomics: Theoretical and experimental considerations." Biotechnology Progress **15**(3): 312-318.

Hawkes, F. R., R. Dinsdale, D. L. Hawkes and I. Hussy (2002). "Sustainable fermentative hydrogen production: Challenges for process optimisation." International Journal of Hydrogen Energy **27**(11-12): 1339-1347.

Heijnen, J. J. (2005). "Approximative kinetic formats used in metabolic network modeling." Biotechnology and Bioengineering **91**(5): 534-545.

Henry, C. S., L. J. Broadbelt and V. Hatzimanikatis (2007). "Thermodynamics-based metabolic flux analysis." Biophys J **92**(5): 1792-805.

Herrgård, M. J., M. W. Covert and B. O. Palsson (2004). "Reconstruction of microbial transcriptional regulatory networks." Current Opinion in Biotechnology **15**(1): 70-77.

Huber, R., T. A. Langworthy, H. Konig, M. Thomm, C. R. Woese, U. B. Sleytr and K. O. Stetter (1986). "*Thermotoga maritima* sp-nov represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C." Archives of Microbiology **144**(4): 324-333.

Iannotti, E. L., Kafkewit.D, M. J. Wolin and M. P. Bryant (1973). "Glucose fermentation products of *Ruminococcus albus* grown in continuous culture with *Vibrio succinogenes* - changes caused by interspecies transfer of H_2." Journal of Bacteriology **114**(3): 1231-1240.

Jannasch, H. W., R. Huber, S. Belkin and K. O. Stetter (1988). "*Thermotoga neapolitana* sp-nov of the extremely thermophilic, eubacterial genus *Thermotoga*." Archives of Microbiology **150**(1): 103-104.

Janssen, P. H. and H. W. Morgan (1992). "Glucose catabolism by *Spirochaeta thermophila* Ri 19.B1." Journal of Bacteriology **174**(8): 2449-2453.

Janssen, P. H. and H. W. Morgan (1992). "Heterotrophic sulfur reduction by *Thermotoga* sp strain Fjss3.B1." Fems Microbiology Letters **96**(2-3): 213-218.

Jenney, F. E. and M. W. W. Adams (2008). "Hydrogenases of the model hyperthermophiles." Incredible Anaerobes: From Physiology to Genomics to Fuels **1125**: 252-266.

Juszczak, A., S. Aono and M. W. W. Adams (1991). "The Extremely Thermophilic Eubacterium, *Thermotoga maritima*, Contains a Novel Iron-Hydrogenase Whose Cellular-Activity Is Dependent Upon Tungsten." Journal of Biological Chemistry **266**(21): 13834-13841.

Kabir, M. M., P. Y. Ho and K. Shimizu (2005). "Effect of ldhA gene deletion on the metabolism of *Escherichia coli* based on gene expression, enzyme activities, intracellular metabolite concentrations, and metabolic flux distribution." Biochemical Engineering Journal **26**(1): 1-11.

Kadar, Z., T. De Vrije, M. A. W. Budde, Z. Szengyel, K. Reczey and P. A. M. Claassen (2003). "Hydrogen production from paper sludge hydrolysate." Applied Biochemistry and Biotechnology **105**: 557-566.

Kadar, Z., T. De Vrijek, G. E. van Noorden, M. A. W. Budde, Z. Szengyel, K. Reczey and P. A. M. Claassen (2004). "Yields from glucose, xylose, and paper sludge hydrolysate during hydrogen production by the extreme thermophile *Caldicellulosiruptor saccharolyticus*." Applied Biochemistry and Biotechnology **113-16**: 497-508.

Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe and M. Hirakawa (2010). "KEGG for representation and analysis of molecular networks involving diseases and drugs." Nucleic Acids Res **38**(Database issue): D355-60.

Kapdan, I. K. and F. Kargi (2006). "Bio-hydrogen production from waste materials." Enzyme and Microbial Technology **38**(5): 569-582.

Käslin, S. A., S. E. Childers and K. M. Noll (1998). "Membrane-associated redox activities in *Thermotoga neapolitana*." Archives of Microbiology **170**(4): 297-303.

Kataoka, N., A. Miya and K. Kiriyama (1997). "Studies on hydrogen production by continuous culture system of hydrogen-producing anaerobic bacteria." Water Science and Technology **36**(6-7): 41-47.

Kawaguchi, H., K. Hashimoto, K. Hirata and K. Miyamoto (2001). "H_2 production from algal biomass by a mixed culture of *Rhodobium marinum* A-501 and *Lactobacillus amylovorus*." Journal of Bioscience and Bioengineering **91**(3): 277-282.

Keasling, J. D., J. Benemann, J. Pramanik, T. A. Carrier, K. L. Jones and S. J. Van Dien (1998). A toolkit for metabolic engineering of bacteria: Application to hydrogen production. BioHydrogen. O. R. Zaborsky. New York, Plenum Press**:** 87-97.

Keating, S. M., B. J. Bornstein, A. Finney and M. Hucka (2006). "SBMLToolbox: An SBML toolbox for MATLAB users." Bioinformatics **22**(10): 1275-1277.

Kelly, R. M. and M. W. W. Adams (1994). "Metabolism in hyperthermophilic microorganisms." Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology **66**(1-3): 247-270.

Kengen, S. W. M. and A. J. M. Stams (1994). "Formation of L-alanine as a reduced end-product in carbohydrate fermentation by the hyperthermophilic archaeon *Pyrococcus furiosus*." Archives of Microbiology **161**(2): 168-175.

King, M. R., D. A. Yernool, D. E. Eveleigh and B. M. Chassy (1998). "Thermostable α-galactosidase from *Thermotoga neapolitana*: cloning, sequencing and expression." Fems Microbiology Letters **163**(1): 37-42.

Koku, H., I. Eroglu, U. Gunduz, M. Yucel and L. Turker (2002). "Aspects of the metabolism of hydrogen production by *Rhodobacter sphaeroides*." International Journal of Hydrogen Energy **27**(11-12): 1315-1329.

Kriegshauser, G. and W. Liebl (2000). "Pullulanase from the hyperthermophilic bacterium *Thermotoga maritima*: purification by β-cyclodextrin affinity chromatography." Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **737**(1-2): 245-251.

Kumar, N. and D. Das (2000). "Enhancement of hydrogen production by *Enterobacter cloacae* IIT-BT 08." Process Biochemistry **35**(6): 589-593.

Kumar, N. and D. Das (2001). "Continuous hydrogen production by immobilized *Enterobacter cloacae* IIT-BT 08 using lignocellulosic materials as solid matrices." Enzyme and Microbial Technology **29**(4-5): 280-287.

Kumar, N., A. Ghosh and D. Das (2001). "Redirection of biochemical pathways for the enhancement of H_2 production by *Enterobacter cloacae*." Biotechnology Letters **23**(7): 537-541.

Lay, J. J. (2000). "Modeling and optimization of anaerobic digested sludge converting starch to hydrogen." Biotechnology and Bioengineering **68**(3): 269-278.

Lay, J. J. (2001). "Biohydrogen generation by mesophilic anaerobic fermentation of microcrystalline cellulose." Biotechnology and Bioengineering **74**(4): 280-287.

Lay, J. J., K. S. Fan, J. I. Hwang, J. I. Chang and P. C. Hsu (2005). "Factors affecting hydrogen production from food wastes by Clostridium-rich composts." Journal of Environmental Engineering-Asce **131**(4): 595-602.

Le Fourn, C., M. L. Fardeau, B. Ollivier, E. Lojou and A. Dolla (2008). "The hyperthermophilic anaerobe *Thermotoga maritima* is able to cope with limited amount of oxygen: insights into its defence strategies." Environmental Microbiology **10**(7): 1877-1887.

Lee, D. S., H. Burd, J. Liu, E. Almaas, O. Wiest, A. L. Barabasi, Z. N. Oltvai and V. Kapatral (2009). "Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets." Journal of Bacteriology **191**(12): 4015-4024.

Lee, M. J. and S. H. Zinder (1988). "Hydrogen partial pressures in a thermophilic acetate-oxidizing methanogenic coculture." Applied and Environmental Microbiology **54**(6): 1457-1461.

Lever, M. (1972). "New reaction for colorimetric determination of carbohydrates." Analytical Biochemistry **47**(1): 273-279.

Levin, D. B., R. Islam, N. Cicek and R. Sparling (2006). "Hydrogen production by *Clostridium thermocellum* 27405 from cellulosic biomass substrates." International Journal of Hydrogen Energy **31**(11): 1496-1503.

Levin, D. B., L. Pitt and M. Love (2004). "Biohydrogen production: prospects and limitations to practical application." International Journal of Hydrogen Energy **29**(2): 173-185.

Liebl, W. (2001). "Cellulolytic enzymes from Thermotoga species." Hyperthermophilic Enzymes, Pt A **330**: 290-300.

Liebl, W., I. Stemplinger and P. Ruile (1997). "Properties and gene structure of the *Thermotoga maritima* α-amylase AmyA, a putative lipoprotein of a hyperthermophilic bacterium." Journal of Bacteriology **179**(3): 941-948.

Liebl, W. G., C. Winterhalter, W. Baumeister, M. Armbrech and M. Valdez (2008). "Xylanase attachment to the cell wall of the hyperthermophilic bacterium *Thermotoga maritima*." Journal of Bacteriology **190**(4): 1350-1358.

Lim, S. K., J. S. Kim, S. H. Cha, B. C. Park, D. S. Lee, H. S. Tae, S. J. Kim, J. J. Kim, K. J. Park and S. Y. Lee (2009). *Thermotoga neapolitana* DSM 4359, complete genome. Accession Number NC_011978, NCBI.

Lin, C. Y. and C. H. Lay (2004). "Carbon/nitrogen-ratio effect on fermentative hydrogen production by mixed microflora." International Journal of Hydrogen Energy **29**(1): 41-45.

Lin, C. Y. and C. H. Lay (2004). "Effects of carbonate and phosphate concentrations on hydrogen production using anaerobic sewage sludge microflora." International Journal of Hydrogen Energy **29**(3): 275-281.

Lo, Y.-C., W.-M. Chen, C.-H. Hung, S.-D. Chen and J.-S. Chang (2008). "Dark H_2 fermentation from sucrose and xylose using H_2-producing indigenous bacteria: Feasibility and kinetic studies." Water Research **42**(4-5): 827-842.

Logan, B. and S. Grot (2006). A bio-electrochemically assisted microbial reactor (BEAMR) that generates hydrogen gas. United States.

Lopes Pinto, F. A., O. Troshina and P. Lindblad (2002). "A brief look at three decades of research on cyanobacterial hydrogen evolution." International Journal of Hydrogen Energy **27**(11-12): 1209-1215.

Luedeking, R. and E. L. Piret (1959). "Transient and steady states in continuous fermentation - Theory and experiment." Journal of Biochemical and Microbiological Technology and Engineering **1**(4): 431-459.

Lunina, N. A., O. V. Berezina, B. Veith, V. V. Zverlov, I. P. Vorobjeva, L. A. Chekanovskaya, I. S. Khromov, C. Raasch, W. Liebl and G. A. Velikodvorskaya (2003). "A cluster of *Thermotoga neapolitana* genes involved in the degradation of starch and maltodextrins: The expression of the aglB and aglA genes in *E. coli* and the properties of the recombinant enzymes." Molecular Biology **37**(5): 686-694.

Madigan, M. T., J. M. Martinko and J. Parker (2003). Brock Biology of Microorganisms. Upper Saddle River, NJ, Prentice Hall.

Mahadevan, R., D. R. Bond, J. E. Butler, A. Esteve-Nunez, M. V. Coppi, B. O. Palsson, C. H. Schilling and D. R. Lovley (2006). "Characterization of metabolism in the Fe(III)-reducing organism *Geobacter sulfurreducens* by constraint-based modeling." Applied and Environmental Microbiology **72**(2): 1558-1568.

Makhorin, A. (2008). Gnu linear programming kit (GLPK). Moscow.

Mamo, G., R. Hatti-Kaul and B. Mattiasson (2007). "Fusion of carbohydrate binding modules from *Thermotoga neapolitana* with a family 10 xylanase from *Bacillus halodurans* S7." Extremophiles **11**(1): 169-177.

Manish, S., K. V. Venkatesh and R. Banerjee (2007). "Metabolic flux analysis of biological hydrogen production by *Escherichia coli*." International Journal of Hydrogen Energy **32**(16): 3820-3830.

Markov, S. A., A. D. Thomas, M. J. Bazin and D. O. Hall (1997). "Photoproduction of hydrogen by cyanobacteria under partial vacuum in batch culture or in a photobioreactor." International Journal of Hydrogen Energy **22**(5): 521-524.

Markowitz, V. M., I. M. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, I. Anderson, A. Lykidis, K. Mavromatis, N. N. Ivanova and N. C. Kyrpides (2010). "The integrated microbial genomes system: An expanding comparative analysis resource." Nucleic Acids Res **38**(Database issue): D382-90.

Mashego, M. R., K. Rumbold, M. De Mey, E. Vandamme, W. Soetaert and J. J. Heijnen (2007). "Microbial metabolomics: Past, present and future methodologies." Biotechnology Letters **29**(1): 1-16.

McCarthy, J. K., A. Uzelac, D. F. Davis and D. E. Eveleigh (2004). "Improved catalytic efficiency and active site modification of 1,4-β-D-glucan glucohydrolase A from *Thermotoga neapolitana* by directed evolution." Journal of Biological Chemistry **279**(12): 11495-11502.

Minnan, L., H. Jinli, W. Xiaobin, X. Huijuan, C. Jinzao, L. Chuannan, Z. Fengzhang and X. Liangshu (2005). "Isolation and characterization of a high H_2-producing strain *Klebsiella oxytoca* HP1 from a hot spring." Research in Microbiology **156**(1): 76-81.

Mongodin, E. F., I. R. Hance, R. T. DeBoy, S. R. Gill, S. Daugherty, R. Huber, C. M. Fraser, K. Stetter and K. E. Nelson (2005). "Gene transfer and genome plasticity in *Thermotoga maritima*, a model hyperthermophilic species." Journal of Bacteriology **187**(14): 4935-4944.

Montgomery, D. C. (2005). Design and analysis of experiments. Hoboken, NJ, Wiley.

Morimoto, M., M. Atsuko, A. A. Y. Atif, M. A. Ngan, A. Fakhru'l-Razi, S. E. Iyuke and A. M. Bakir (2004). "Biological production of hydrogen from glucose by natural anaerobic microflora." International Journal of Hydrogen Energy **29**(7): 709-713.

Mu, Y., G. Wang and H. Q. Yu (2006). "Kinetic modeling of batch hydrogen production process by mixed anaerobic cultures." Bioresource Technology **97**(11): 1302-1307.

Mu, Y., H. Q. Yu and G. Wang (2007). "A kinetic approach to anaerobic hydrogen-producing process." Water Research **41**(5): 1152-1160.

Munro, S. A., S. H. Zinder and L. P. Walker (2009). "The fermentation stoichiometry of *Thermotoga neapolitana* and influence of temperature, oxygen, and pH on hydrogen production." Biotechnology Progress **25**(4): 1035-1042.

Nanavati, D., K. M. Noll and A. H. Romano (2002). "Periplasmic maltose- and glucose-binding protein activities in cell-free extracts of *Thermotoga maritima*." Microbiology-Sgm **148**: 3531-3537.

Nanavati, D. A., K. Thirangoon and K. A. Noll (2006). "Several archaeal homologs of putative oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars." Applied and Environmental Microbiology **72**(2): 1336-1345.

Nanavati, D. M., T. N. Nguyen and K. M. Noll (2005). "Substrate specificities and expression patterns reflect the evolutionary divergence of maltose ABC transporters in *Thermotoga maritima*." Journal of Bacteriology **187**(6): 2002-2009.

Nath, K. and D. Das (2004). "Improvement of fermentative hydrogen production: Various approaches." Applied Microbiology and Biotechnology **65**(5): 520-529.

Nath, K., M. Muthukumar, A. Kumar and D. Das (2008). "Kinetics of two-stage fermentation process for the production of hydrogen." International Journal of Hydrogen Energy **33**(4): 1195-1203.

Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, L. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter and C. M. Fraser (1999). "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*." Nature **399**(6734): 323-329.

Nelson, K. E., J. A. Eisen and C. M. Fraser (2001). "Genome of *Thermotoga maritima* MSB8." Hyperthermophilic Enzymes, Pt A **330**: 169-180.

Nesbø, C. L., M. Dlutek and W. F. Doolittle (2006). "Recombination in thermotoga: Implications for species concepts and biogeography." Genetics **172**(2): 759-769.

Nesbø, C. L., S. L'Haridon, K. O. Stetter and W. F. Doolittle (2001). "Phylogenetic analyses of two "Archaeal" genes in *Thermotoga maritima* reveal multiple transfers between Archaea and Bacteria." Molecular Biology and Evolution **18**(3): 362-375.

Nesbø, C. L., K. E. Nelson and W. F. Doolittle (2002). "Suppressive subtractive hybridization detects extensive genomic diversity in *Thermotoga maritima*." Journal of Bacteriology **184**(16): 4475-4488.

Nguyen, T. A. D., J. P. Kim, M. S. Kim, Y. K. Oh and S. J. Sim (2008). "Optimization of hydrogen production by hyperthermophilic eubacteria, *Thermotoga maritima* and *Thermotoga neapolitana* in batch fermentation." International Journal of Hydrogen Energy **33**(5): 1483-1488.

Nguyen, T. N., K. M. Borges, A. H. Romano and K. M. Noll (2001). "Differential gene expression in *Thermotoga neapolitana* in response to growth substrate." Fems Microbiology Letters **195**(1): 79-83.

Nguyen, T. N., A. D. Ejaz, M. A. Brancieri, A. M. Mikula, K. E. Nelson, S. R. Gill and K. M. Noll (2004). "Whole-genome expression profiling of *Thermotoga maritima* in response to growth on sugars in a chemostat." Journal of Bacteriology **186**(14): 4824-4828.

Nielsen, J., K. Nikolajsen and J. Villadsen (1991). "Structured modeling of a microbial system 1. A theoretical-study of lactic-acid fermentation." Biotechnology and Bioengineering **38**(1): 1-10.

Noike, T. and O. Mizuno (2000). "Hydrogen fermentation of organic municipal wastes." Water Science and Technology **42**(12): 155-162.

Noll, K. M. and M. Vargas (1997). "Recent advances in genetic analyses of hyperthermophilic Archaea and Bacteria." Archives of Microbiology **168**(2): 73-80.

Oh, S. E. and B. E. Logan (2005). "Hydrogen and electricity production from a food processing wastewater using fermentation and microbial fuel cell technologies." Water Research **39**(19): 4673-4682.

Oh, S. E., S. Van Ginkel and B. E. Logan (2003). "The relative effectiveness of pH control and heat treatment for enhancing biohydrogen gas production." Environmental Science & Technology **37**(22): 5186-5190.

Oh, Y. K., H. J. Kim, S. Park, M. S. Kim and D. D. Y. Ryu (2008). "Metabolic-flux analysis of hydrogen production pathway in *Citrobacter amalonaticus* Y19." International Journal of Hydrogen Energy **33**(5): 1471-1482.

Oh, Y. K., E. H. Seol, J. R. Kim and S. Park (2003). "Fermentative biohydrogen production by a new chemoheterotrophic bacterium *Citrobacter* sp Y19." International Journal of Hydrogen Energy **28**(12): 1353-1359.

Palsson, B. (2006). Systems biology : Properties of reconstructed networks. Cambridge ; New York, Cambridge University Press.

Parker, K. N., S. R. Chhabra, D. Lam, W. Callen, G. D. Duffaud, M. A. Snead, J. M. Short, E. J. Mathur and R. M. Kelly (2001). "Galactomannanases man2 and man5 from *Thermotoga* species: Growth physiology on galactomannans, gene sequence analysis, and biochemical properties of recombinant enzymes." Biotechnology and Bioengineering **75**(3): 322-333.

Patil, K. R., M. Akesson and J. Nielsen (2004). "Use of genome-scale microbial models for metabolic engineering." Current Opinion in Biotechnology **15**(1): 64-69.

Price, N. D., J. L. Reed and B. O. Palsson (2004). "Genome-scale models of microbial cells: Evaluating the consequences of constraints." Nature Reviews Microbiology **2**(11): 886-897.

Pysz, M. A., S. B. Conners, C. I. Montero, K. R. Shockley, M. R. Johnson, D. E. Ward and R. A. Kelly (2004). "Transcriptional analysis of biofilm formation processes in the anaerobic, hyperthermophilic bacterium *Thermotoga maritima*." Applied and Environmental Microbiology **70**(10): 6098-6112.

Pysz, M. A., D. E. Ward, K. R. Shockley, C. I. Montero, S. B. Conners, M. R. Johnson and R. M. Kelly (2004). "Transcriptional analysis of dynamic heat-shock response by the hyperthermophilic bacterium *Thermotoga maritima*." Extremophiles **8**(3): 209-217.

Rainey, F. A., P. H. Janssen, D. J. C. Wild and H. W. Morgan (1991). "Isolation and characterization of an obligately anaerobic, polysaccharolytic, extremely thermophilic member of the genus *Spirochaeta*." Archives of Microbiology **155**(4): 396-401.

Ravot, G., B. Ollivier, M. L. Fardeau, B. K. C. Patel, K. T. Andrews, M. Magot and J. L. Garcia (1996). "L-Alanine production from glucose fermentation by hyperthermophilic members of the domains Bacteria and Archaea: A remnant of an ancestral metabolism?" Applied and Environmental Microbiology **62**(7): 2657-2659.

Rinker, K. D. and R. M. Kelly (2000). "Effect of carbon and nitrogen sources on growth dynamics and exopolysaccharide production for the hyperthermophilic archaeon *Thermococcus litoralis* and bacterium *Thermotoga maritima*." Biotechnology and Bioengineering **69**(5): 537-547.

Rittmann, B. E. and P. L. McCarty (2001). Environmental biotechnology : Principles and applications. Boston, McGraw-Hill.

Roberts, S. B., J. L. Robichaux, A. K. Chavali, P. A. Manque, V. Lee, A. M. Lara, J. A. Papin and G. A. Buck (2009). "Proteomic and network analysis characterize stage-specific metabolism in *Trypanosoma cruzi*." BMC Syst Biol **3**: 52.

Rupprecht, J., B. Hankamer, J. H. Mussgnug, G. Ananyev, C. Dismukes and O. Kruse (2006). "Perspectives and advances of biological H_2 production in microorganisms." Applied Microbiology and Biotechnology **72**(3): 442-449.

Saeki, K. and H. Kumagai (1998). "The rnf gene products in *Rhodobacter capsulatus* play an essential role in nitrogen fixation during anaerobic DMSO-dependent growth in the dark." Archives of Microbiology **169**(5): 464-467.

Saint-Amans, S., L. Girbal, J. Andrade, K. Ahrens and P. Soucaille (2001). "Regulation of carbon and electron flow in *Clostridium butyricum* VPI 3266 grown on glucose-glycerol mixtures." Journal of Bacteriology **183**(5): 1748-1754.

Sauer, U. (2004). "High-throughput phenomics: Experimental methods for mapping fluxomes." Current Opinion in Biotechnology **15**(1): 58-63.

Schäfer, T. and P. Schönheit (1992). "Maltose fermentation to acetate, CO_2 and H_2 in the anaerobic hyperthermophilic archaeon *Pyrococcus furiosus* - evidence for the operation of a novel sugar fermentation pathway." Archives of Microbiology **158**(3): 188-202.

Schilling, C. H., M. W. Covert, I. Famili, G. M. Church, J. S. Edwards and B. O. Palsson (2002). "Genome-scale metabolic model of *Helicobacter pylori* 26695." Journal of Bacteriology **184**(16): 4582-4593.

Schönheit, P. and T. Schäfer (1995). "Metabolism of hyperthermophiles." World Journal of Microbiology & Biotechnology **11**(1): 26-57.

Schröder, C., M. Selig and P. Schönheit (1994). "Glucose fermentation to acetate, CO_2 and H_2 in the anaerobic hyperthermophilic eubacterium *Thermotoga maritima* - involvement of the Embden-Meyerhof pathway." Archives of Microbiology **161**(6): 460-470.

Schumann, J., A. Wrba, R. Jaenicke and K. O. Stetter (1991). "Topographical and enzymatic characterization of amylases from the extremely thermophilic eubacterium *Thermotoga maritima*." FEBS Letters **282**(1): 122-126.

Schut, G. J. and M. W. Adams (2009). "The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: A new perspective on anaerobic hydrogen production." J Bacteriol **191**(13): 4451-7.

Selig, M., K. B. Xavier, H. Santos and P. Schonheit (1997). "Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*." Archives of Microbiology **167**(4): 217-232.

Sen, D. and D. Das (2005). "Multiple parameter optimization for the maximization of hydrogen production by *Enterobacter cloacae* DM11." Journal of Scientific & Industrial Research **64**(12): 984-990.

Senger, R. S. and E. T. Papoutsakis (2008). "Genome-scale model for *Clostridium acetobutylicum*: Part I. Metabolic network resolution and analysis." Biotechnol Bioeng **101**(5): 1036-52.

Shuler, M. L. and F. Kargi (2002). Bioprocess engineering : Basic concepts. Upper Saddle River, NJ, Prentice Hall PTR.

Soutschek, E., J. Winter, F. Schindler and O. Kandler (1984). "*Acetomicrobium flavidum*, gen-nov, sp-nov, a thermophilic, anaerobic bacterium from sewage-sludge, forming acetate, CO_2 and H_2 from glucose." Systematic and Applied Microbiology **5**(3): 377-390.

Stephanopoulos, G., A. A. Aristidou and J. Nielsen (1998). Metabolic engineering : Principles and methodologies. San Diego, Academic Press.

Stoscheck, C. M. (1990). "Increased uniformity in the response of the Coomassie Blue G-Protein assay to different proteins." Analytical Biochemistry **184**(1): 111-116.

Sunna, A., J. Puls and G. Antranikian (1997). "Characterization of the xylanolytic enzyme system of the extreme thermophilic anaerobic bacteria *Thermotoga maritima*, *T.neapolitana*, and *T. thermarum*." Comparative Biochemistry and Physiology a-Physiology **118**(3): 453-461.

Taguchi, F., N. Mizukami, T. S. Taki and K. Hasegawa (1995). "Hydrogen-production from continuous fermentation of xylose during growth of *Clostridium* sp strain No-2." Canadian Journal of Microbiology **41**(6): 536-540.

Tamagnini, P., R. Axelsson, P. Lindberg, F. Oxelfelt, R. Wunschiers and P. Lindblad (2002). "Hydrogenases and hydrogen metabolism of cyanobacteria." Microbiology and Molecular Biology Reviews **66**(1): 1-16.

Teusink, B., J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff and J. L. Snoep (2000). "Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry." European Journal of Biochemistry **267**(17): 5313-5329.

Thauer, R. K., K. Jungermann and K. Decker (1977). "Energy-conservation in chemotropic anaerobic bacteria." Bacteriological Reviews **41**(1): 100-180.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nat Protoc **5**(1): 93-121.

Tosatto, S. C. E., S. Toppo, D. Carbonera, G. M. Giacometti and P. Costantini (2008). "Comparative analysis of [FeFe] hydrogenase from *Thermotogales* indicates the molecular basis of resistance to oxygen inactivation." International Journal of Hydrogen Energy **33**(2): 570-578.

Tsygankov, A., S. Kosourov, M. Seibert and M. L. Ghirardi (2002). "Hydrogen photoproduction under continuous illumination by sulfur-deprived, synchronous *Chlamydomonas reinhardtii* cultures." International Journal of Hydrogen Energy **27**(11-12): 1239-1244.

Tsygankov, A. A., A. S. Fedorov, S. N. Kosourov and K. K. Rao (2002). "Hydrogen production by cyanobacteria in an automated outdoor photobioreactor under aerobic conditions." Biotechnology and Bioengineering **80**(7): 777-783.

Turner, J. A. (2004). "Sustainable hydrogen production." Science **305**(5686): 972-974.

Ueno, Y., S. Haruta, M. Ishii and Y. Igarashi (2001). "Characterization of a microorganism isolated from the effluent of hydrogen fermentation by microflora." Journal of Bioscience and Bioengineering **92**(4): 397-400.

Ueno, Y., T. Kawai, S. Sato, S. Otsuka and M. Morimoto (1995). "Biological production of hydrogen from cellulose by natural anaerobic microflora." Journal of Fermentation and Bioengineering **79**(4): 395-397.

Van Ginkel, S., S. W. Sung and J. J. Lay (2001). "Biohydrogen production as a function of pH and substrate concentration." Environmental Science & Technology **35**(24): 4726-4730.

Van Ginkel, S. W. and B. Logan (2005). "Increased biological hydrogen production with reduced organic loading." Water Research **39**(16): 3819-3826.

van Groenestijn, J. W., J. H. O. Hazewinkel, M. Nienoord and P. J. T. Bussmann (2002). "Energy aspects of biological hydrogen production in high rate bioreactors operated in the thermophilic temperature range." International Journal of Hydrogen Energy **27**(11-12): 1141-1147.

van Niel, E. W. J., P. A. M. Claassen and A. J. M. Stams (2003). "Substrate and product inhibition of hydrogen production by the extreme thermophile, *Caldicellulosiruptor saccharolyticus*." Biotechnology and Bioengineering **81**(3): 255-262.

Van Ooteghem, S. (2005). Process for generation of hydrogen gas from various feedstocks using thermophilic bacteria.. United States, The United States of America as represented by the United States Department of Energy.

Van Ooteghem, S. A., S. K. Beer and P. C. Yue (2002). "Hydrogen production by the thermophilic bacterium *Thermotoga neapolitana*." Applied Biochemistry and Biotechnology **98**: 177-189.

Van Ooteghem, S. A., A. Jones, D. van der Lelie, B. Dong and D. Mahajan (2004). "H_2 production and carbon utilization by *Thermotoga neapolitana* under anaerobic and microaerobic growth conditions." Biotechnology Letters **26**(15): 1223-1232.

Vargas, M. and K. M. Noll (1994). "Isolation of auxotrophic and antimetabolite-resistant mutants of the hyperthermophilic bacterium *Thermotoga neapolitana*." Archives of Microbiology **162**(5): 357-361.

Vargas, M. and K. M. Noll (1996). "Catabolite repression in the hyperthermophilic bacterium *Thermotoga neapolitana* is independent of cAMP." Microbiology-Uk **142**: 139-144.

Varma, A. and B. O. Palsson (1994). "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product decretion in wild-type *Escherichia coli* W3110." Applied and Environmental Microbiology **60**(10): 3724-3731.

Veith, B., V. V. Zverlov, N. A. Lunina, O. V. Berezina, C. Raasch, G. A. Velikodvorskaya and W. Liebl (2003). "Comparative analysis of the recombinant α-glucosidases from the *Thermotoga neapolitana* and *Thermotoga maritima* maltodextrin utilization gene clusters." Biocatalysis and Biotransformation **21**(4-5): 147-158.

Velikodvorskaya, T. V., I. Y. Volkov, V. T. Vasilevko, V. V. Zverlov and E. S. Piruzian (1997). "Purification and some properties of *Thermotoga neapolitana* thermostable xylanase B expressed in *E. coli* cells." Biochemistry-Moscow **62**(1): 66-70.

Verhagen, M., T. O'Rourke and M. W. W. Adams (1999). "The hyperthermophilic bacterium, *Thermotoga maritima*, contains an unusually complex iron-hydrogenase: amino acid sequence analyses versus biochemical characterization." Biochimica Et Biophysica Acta-Bioenergetics **1412**(3): 212-229.

Verhagen, M. F. J. M. and M. W. W. Adams (2001). "Fe-only hydrogenase from *Thermotoga maritima*." Hyperthermophilic Enzymes, Pt B **331**: 216-226.

Vining, G. G. (1998). Statistical methods for engineers. Pacific Grove, CA., Duxbury Press.

Wang, D. I. C., C. I. Cooney, A. L. Demain, P. Dunnill, A. E. Humphrey and M. D. Lilly (1979). Fermentation and enzyme technology. New York, John Wiley & Sons.

Westermann, P., B. Jorgensen, L. Lange, B. K. Ahring and C. H. Christensen (2007). "Maximizing renewable hydrogen production from biomass in a bio/catalytic refinery." International Journal of Hydrogen Energy **32**(17): 4135-4141.

Whang, L. M., C. J. Hsiao and S. S. Cheng (2006). "A dual-substrate steady-state model for biological hydrogen production in an anaerobic hydrogen fermentation process." Biotechnology and Bioengineering **95**(3): 492-500.

Wilson, D. B. (2008). "Three microbial strategies for plant cell wall degradation." Incredible Anaerobes: From Physiology to Genomics to Fuels **1125**: 289-297.

Woodward, J., N. I. Heyer, J. P. Getty, H. M. O'Neill, E. Pinkhassik and B. R. Evans (2002). Efficient hydrogen production using enzymes of the pentose phosphate pathway. Proceedings of the 2002 U.S. DOE Hydrogen Program Review, Washington, D.C., US Department of Energy.

Woodward, J., M. Orr, K. Cordray and E. Greenbaum (2000). "Biotechnology - enzymatic production of biohydrogen." Nature **405**(6790): 1014-1015.

Yernool, D. A., J. K. McCarthy, D. E. Eveleigh and J. D. Bok (2000). "Cloning and characterization of the glucooligosaccharide catabolic pathway β-glucan glucohydrolase and cellobiose phosphorylase in the marine hyperthermophile *Thermotoga neapolitana*." Journal of Bacteriology **182**(18): 5172-5179.

Yernool, D. A., R. F. Sullivan, S. Y. Rani and D. E. Eveleigh (2002). Molecular and functional characterization of a xylan catabolic cluster and associated ABC transporter system from the hyperthermophile *Thermotoga neapolitana*.

Yokoi, H., R. Maki, J. Hirose and S. Hayashi (2002). "Microbial production of hydrogen from starch-manufacturing wastes." Biomass & Bioenergy **22**(5): 389-395.

Yokoi, H., S. Mori, J. Hirose, S. Hayashi and Y. Takasaki (1998). "H_2 production from starch by a mixed culture of *Clostridium butyricum* and *Rhodobacter* sp. M-19." Biotechnology Letters **20**(9): 895-899.

Yokoi, H., T. Tokushige, J. Hirose, S. Hayashi and Y. Takasaki (1998). "H_2 production from starch by a mixed culture of *Clostridium butyricum* and *Enterobacter aerogenes*." Biotechnology Letters **20**(2): 143-147.

Yu, J. S., M. Vargas, C. Mityas and K. M. Noll (2001). "Liposome-mediated DNA uptake and transient expression in *Thermotoga*." Extremophiles **5**(1): 53-60.

Zhang, T., H. Liu and H. H. P. Fang (2003). "Biohydrogen production from starch in wastewater under thermophilic condition." Journal of Environmental Management **69**(2): 149-156.

Zhang, Y., I. Thiele, D. Weekes, Z. W. Li, L. Jaroszewski, K. Ginalski, A. M. Deacon, J. Wooley, S. A. Lesley, I. A. Wilson, B. Palsson, A. Osterman and A. Godzik (2009). "Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*." Science **325**(5947): 1544-1549.

Zhaxybayeva, O., K. S. Swithers, P. Lapierre, G. P. Fournier, D. M. Bickhart, R. T. DeBoy, K. E. Nelson, C. L. Nesbo, W. F. Doolittle, J. P. Gogarten and K. M. Noll (2009). "On the chimeric nature, thermophilic origin, and phylogenetic placement of the *Thermotogales*." Proc Natl Acad Sci U S A **106**(14): 5865-70.

Zheng, X. J. and H. Q. Yu (2005). "Inhibitory effects of butyrate on biological hydrogen production with mixed anaerobic cultures." Journal of Environmental Management **74**(1): 65-70.

Zinder, S. H. (2008). S. A. Munro**:** Personal Communication.

Zverlov, V., K. Piotukh, O. Dakhova, G. Velikodvorskaya and R. Borriss (1996). "The multidomain xylanase A of the hyperthermophilic bacterium *Thermotoga neapolitana* is extremely thermoresistant." Applied Microbiology and Biotechnology **45**(1-2): 245-247.

Zverlov, V. V., I. Y. Volkov, T. V. Velikodvorskaya and W. H. Schwarz (1997). "Highly thermostable endo-1,3-β-glucanase (laminarinase) LamA from *Thermotoga neapolitana*: Nucleotide sequence of the gene and characterization of the recombinant gene product." Microbiology-Uk **143**: 1701-1708.

Zverlov, V. V., I. Y. Volkov, T. V. Velikodvorskaya and W. H. Schwarz (1997). "*Thermotoga neapolitana* bglB gene, upstream of lamA, encodes a highly thermostable β-glucosidase that is a laminaribiase." Microbiology-Uk **143**: 3537-3542.